

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2001, 09:54:32 ; Search time 89.59 Seconds
(without alignments)
332.928 Million cell updates/sec

Title: US-09-668-482-2

Perfect score: 2563
Sequence: 1 MGXYTMWFLCTIVLPVL.....GPTIYVDNLPTFTSVRN 492

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

A_Geneseq.0601:*
1: /SIDSL/gcgdata/geneseq/geneseq/AA1980.DAT:*
2: /SIDSL/gcgdata/geneseq/geneseq/AA1981.DAT:*
3: /SIDSL/gcgdata/geneseq/geneseq/AA1982.DAT:*
4: /SIDSL/gcgdata/geneseq/geneseq/AA1983.DAT:*
5: /SIDSL/gcgdata/geneseq/geneseq/AA1984.DAT:*
6: /SIDSL/gcgdata/geneseq/geneseq/AA1985.DAT:*
7: /SIDSL/gcgdata/geneseq/geneseq/AA1986.DAT:*
8: /SIDSL/gcgdata/geneseq/geneseq/AA1987.DAT:*
9: /SIDSL/gcgdata/geneseq/geneseq/AA1988.DAT:*
10: /SIDSL/gcgdata/geneseq/geneseq/AA1989.DAT:*
11: /SIDSL/gcgdata/geneseq/geneseq/AA1990.DAT:*
12: /SIDSL/gcgdata/geneseq/geneseq/AA1991.DAT:*
13: /SIDSL/gcgdata/geneseq/geneseq/AA1992.DAT:*
14: /SIDSL/gcgdata/geneseq/geneseq/AA1993.DAT:*
15: /SIDSL/gcgdata/geneseq/geneseq/AA1994.DAT:*
16: /SIDSL/gcgdata/geneseq/geneseq/AA1995.DAT:*
17: /SIDSL/gcgdata/geneseq/geneseq/AA1996.DAT:*
18: /SIDSL/gcgdata/geneseq/geneseq/AA1997.DAT:*
19: /SIDSL/gcgdata/geneseq/geneseq/AA1998.DAT:*
20: /SIDSL/gcgdata/geneseq/geneseq/AA1999.DAT:*
21: /SIDSL/gcgdata/geneseq/geneseq/AA2000.DAT:*
22: /SIDSL/gcgdata/geneseq/geneseq/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2563	100.0	492	AAW37733	Cytochrome zp450RA
2	2563	100.0	492	AAW44159	zebrafish retinoid
3	1738	67.8	497	AAW37734	Human cytochrome P
4	1738	67.8	497	AAW44160	Human retinoid met
5	1734	67.7	497	AAW37735	Cytochrome P450RAI
6	1734	67.7	497	AAW44161	Mouse retinoid met
7	752.5	29.4	216	AAW12489	Human PSEC64 prote
8	552.5	21.6	472	AAW27153	Arabidopsis thalia
9	552.5	21.6	472	AAW44571	Arabidopsis thalia
10	552.5	21.6	472	AAW45022	Arabidopsis thalia
11	552.5	21.6	491	AAW45021	Arabidopsis thalia

12	552.5	21.6	492	21	AAW44570	Arabidopsis thalia
13	536	20.9	444	21	AAW44572	Arabidopsis thalia
14	536	20.9	444	21	AAW45023	Arabidopsis thalia
15	503	19.6	513	21	AAW07921	A cytochrome P450
16	481.5	18.8	461	21	AAW46490	Arabidopsis thalia
17	481.5	18.8	462	21	AAW46489	Arabidopsis thalia
18	481.5	18.8	465	21	AAW46491	Arabidopsis thalia
19	475.5	18.6	461	21	AAW11836	Arabidopsis thalia
20	475.5	18.6	462	21	AAW11835	Arabidopsis thalia
21	475.5	18.6	465	21	AAW11834	Arabidopsis thalia
22	470	18.3	468	21	AAW20783	Arabidopsis thalia
23	425.5	16.6	430	21	AAW20784	Human ORFX ORF321
24	397	15.5	163	21	AAW40557	Arabidopsis thalia
25	394.5	15.4	489	21	AAW30048	Arabidopsis thalia
26	394	15.4	481	21	AAW30049	Arabidopsis thalia
27	366.5	14.3	433	21	AAW30050	Arabidopsis thalia
28	366	14.3	475	21	AAW07676	Amino acid sequenc
29	343	13.4	511	20	AAW23341	A P450-2 protein I
30	343	13.4	511	20	AAW23341	Sweetgum coniferyl
31	342	13.3	471	21	AAW19694	Arabidopsis thalia
32	342	13.3	479	21	AAW23013	Arabidopsis thalia
33	338.5	13.2	490	17	AAW89862	Cytochrome P450 2C
34	339.5	13.2	490	17	AAW89865	Cytochrome P450 2C
35	339.5	13.2	490	19	AAW64070	Human cytochrome P
36	339.5	13.2	490	19	AAW64073	Human cytochrome P
37	339.5	13.2	490	20	AAW04127	Mammalian cytochro
38	339	13.2	469	21	AAW23014	Arabidopsis thalia
39	333	13.0	388	21	AAW23075	Human cytochrome P
40	332	13.0	490	16	AAW2361	Human cytochrome P
41	332	13.0	490	17	AAW93168	Human cytochrome P
42	332	13.0	490	17	AAW81465	Human derived cyto
43	332	13.0	513	22	AAW31008	Amino acid sequenc
44	332	13.0	513	22	AAW48181	B. napus PSH polyP
45	330.5	12.9	520	18	AAW26640	Arabidopsis thalia

ALIGNMENTS

RESULT 1
AAW37733 standard; Protein; 492 AA.
XX
AC AAW37733;
XX
DT 07-JUL-1998 (first entry)
XX
DE Cytochrome zp450RAI protein.
XX
XX Retinoid regulated gene; cytochrome P450 gene; enzyme;
KW oxidative metabolism; P450RAI; retinoid acid; RA; promoter.
XX
OS Danio rerio.
XX
XX WO9749832-A2.
XX
XX 31-DEC-1997.
XX
XX 23-JUN-1997; 97WO-CA00488.
XX
XX 01-OCT-1996; 96US-0724466.
XX
XX 21-JUN-1996; 96US-0667546.
XX
XX (TOOH) UNIV QUEENS KINGSTON.
XX
XX Petkovich PW;
XX
XX WPI: 1998-077193/07.
XX
XX N-PSDB; AAW09251.
XX
XX Identifying DNA encoding inducible or suppressible cytochrome P450 -
XX by screening for drugs which reduce the catabolism of retinoid acid,
XX useful in cancer chemotherapy and the treatment of acne and

PT psoriasis
 XX
 PS
 XX
 XX
 CC Example 1; Pages 53-55; 113pp; English.
 CC This is the amino acid for cytochrome zp450RAI of the zebra fish.
 CC Its expression is dependent on the presence of retinoic acid (RA).
 CC The retinoid-regulated genes such as the inducible cytochrome P450RAI
 CC gene specifically metabolises a derivative of the RA. The cytochrome
 CC P450 gene in general produces enzymes involved in the oxidative
 CC metabolism of endogenous and exogenous compounds. The cytochrome
 CC P450 nucleotide sequence can be used to induce or suppress the
 CC expression of its protein. P450RAI is highly induced by RA in cell
 CC lines and tissues. This allows for development of a drug screen using
 CC promoters and nucleotide sequences to identify drugs which are useful
 CC for reducing the catabolism of RA.
 CC
 XX
 SQ Sequence 492 AA:
 Query Match 100.0%; Score 2563; DB 19; Length 492;
 Best Local Similarity 100.0%; Pred. No. 5.3e-227;
 Matches 492; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MGLYTLMTFTCTIVLPVLLFLAAVKLMEMLMIRRYDNCSPPLPPTMGLPFTGETIQL 60
 DB 1 mgllytlmvtftctivlpvllflaavkllwemlmirrvdncsrspplppgtmglpfiftgltql 60
 QY 61 ILORRRFLMKRQKYGCIYKTHLFGNPTVRVAGADNVRQILLEGHKLVSVOVPASVRTIL 120
 DB 61 ilgrkflrmkrqkgyckythlfgnptvrmgdnvrqillegkhklysvqwpasvrtll 120
 QY 121 GSDTLSNVHGVQHNKKKRAIMRFSRDALHRYIPVIOEVKSAIOEWLOKDSCVLYPEM 180
 DB 121 gsdtlsnvhgvqhnkkkaimrfsrdalehyipviqgevksaioewlqkds cvlypem 180
 QY 181 KKLMEIRIMRILGPEPQIKTDEQELVEAFEMIKNLSPIIDVPFSGLYGLRARNFI 240
 DB 181 kklmriamrlllgfepqiktdelaveaemknlsfpidvpfsglytglarnfi 240
 QY 241 HSKIEENIRKKIQQDDNENQKYKDALQLIENSRRSDPEFSIQAMKEAATELLFGHET 300
 DB 241 hskleeniirkkiqddnneqkykdalqlilensrrsdepfslqamkeate lllfgghet 300
 QY 301 TASTATSLVMEGLNTEVYVQKREVOEKVEMGMTTPGKLSMELLDQLKTYGCVIKETL 360
 DB 301 tascatslvmlglnteveyvqkreveqkvmgmttpgklsmelldqlkygcviketl 360
 QY 361 RINPVPQGFVALKTEFLNGVQIPKGNVYISICDTHADVADVFENKEEFQERFMSKGL 420
 DB 361 rinppvpqgfvalktfelnvgqipkgwnvysicdthadvadvfnpkeefqerfmskyl 420
 QY 421 EDGSRFNTIPFGGSRMCGVKEPAKVLKIFLVLTQHCHNMTLSNGPPTMTKGTPIYVPD 480
 DB 421 edgsrftnypfggsrcmvgkefakvlkiflvltqhcnmllsnppmtkgtpiyvpd 480
 QY 481 NLPTKFTSYVFN 492
 DB 481 nlptkftsyvfn 492
 RESULT 2
 AAM44159 standard; Protein: 492 AA.
 AC AAM44159;
 XX
 XX 22-JUN-1998 (first entry)
 DE zebrafish retinoid metabolising protein zp450RAI.
 KW Retinoid metabolising protein; P450RAI; retinoid oxidase;
 KW retinoic acid; zebrafish; inhibitor; antibody; cancer;
 KW actinic keratosis; oral leukoplakia; head tumour; neck tumour;

KW non-small cell lung carcinoma; basal cell carcinoma;
 KW acute promyelocytic leukaemia; skin cancer; acne; psoriasis;
 KW ichthyosis; therapy; diagnosis; screening.
 XX
 XX Danio rerio.
 OS
 XX
 XX MO9749815-A1.
 PN
 XX
 PD 31-DEC-1997.
 PF
 XX 23-JUN-1997; 97WO-CA00440.
 XX
 PR 01-OCT-1996; 96US-0724466.
 PR 21-JUN-1996; 96US-0667546.
 XX
 PA (TOOH) UNITV QUEENS KINGSTON.
 XX
 PI Beckett BR, Jones G, Petkovich PM, White JA;
 DR WPI: 1998-077178/07.
 DR N-PSDB: AAV12203.
 XX
 PT Retinoid metabolising protein - useful to develop products to treat,
 PT e.g. cancer, actinic keratosis, oral leukoplakia, acne, psoriasis or
 PT ichthyosis
 PS Claim 1; Page 50-51; 110pp; English.
 CC This protein comprises a novel zebrafish retinoid metabolising
 CC protein, designated zp450RAI. Its amino acid sequence was deduced
 CC from a cDNA clone (see AAV12203) isolated from a 6-18 hr embryo
 CC library. It includes a haem-binding motif characteristic of
 CC cytochrome P450 proteins. zp450RAI is a retinoid oxidase that has
 CC the ability to hydroxylate retinoic acid at the 4 position of the
 CC beta ionone ring, and is inducible in epithelial cells exposed to
 CC retinoic acid. Zebrafish, human and mouse P450RAIs (see AAM44159-61)
 CC are claimed. They can be expressed in host cells and used to
 CC metabolize retinoic acid in an organism or cell, in drug screening,
 CC and to raise antibodies useful for inhibiting retinoic acid
 CC hydroxylation for the treatment of cancer, actinic keratosis, oral
 CC leukoplakia, secondary tumours of the head and/or neck, non-small
 CC cell lung carcinomas, basal cell carcinomas, acute promyelocytic
 CC leukaemia, skin cancer, and premalignancy associated with actinic
 CC keratosis, acne, psoriasis and/or ichthyosis.
 CC
 XX
 SQ Sequence 492 AA:
 Query Match 100.0%; Score 2563; DB 19; Length 492;
 Best Local Similarity 100.0%; Pred. No. 5.3e-227;
 Matches 492; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MGLYTLMTFTCTIVLPVLLFLAAVKLMEMLMIRRYDNCSPPLPPTMGLPFTGETIQL 60
 DB 1 mgllytlmvtftctivlpvllflaavkllwemlmirrvdncsrspplppgtmglpfiftgltql 60
 QY 61 ILORRRFLMKRQKYGCIYKTHLFGNPTVRVAGADNVRQILLEGHKLVSVOVPASVRTIL 120
 DB 61 ilgrkflrmkrqkgyckythlfgnptvrmgdnvrqillegkhklysvqwpasvrtll 120
 QY 121 GSDTLSNVHGVQHNKKKRAIMRFSRDALHRYIPVIOEVKSAIOEWLOKDSCVLYPEM 180
 DB 121 gsdtlsnvhgvqhnkkkaimrfsrdalehyipviqgevksaioewlqkds cvlypem 180
 QY 181 KKLMEIRIMRILGPEPQIKTDEQELVEAFEMIKNLSPIIDVPFSGLYGLRARNFI 240
 DB 181 kklmriamrlllgfepqiktdelaveaemknlsfpidvpfsglytglarnfi 240
 QY 241 HSKIEENIRKKIQQDDNENQKYKDALQLIENSRRSDPEFSIQAMKEAATELLFGHET 300
 DB 241 hskleeniirkkiqddnneqkykdalqlilensrrsdepfslqamkeate lllfgghet 300
 QY 301 TASTATSLVMEGLNTEVYVQKREVOEKVEMGMTTPGKLSMELLDQLKTYGCVIKETL 360

```

|||||
Db 301 testatslmfvgintevvqkveevqkvegmymtpgkqjlsmeilqklkytgcvtketl 360
QY 361 RINPVPVGGFRVALKTFELNGYQIPKGMNVIYSICDTHVDADVFPNKEEPQPERFNSKGL 420
Db 361 RINPVPVGGFRVALKTFELNGYQIPKGMNVIYSICDTHVDADVFPNKEEPQPERFNSKGL 420
QY 421 EGGSRPNYIPFGGSRMCGVEKFAVLKIFLVELTQHNMILNSGPPMTKGTPTIYPVD 480
Db 421 eggsrpnypfggsrmevgekefakvllkiflveltnhnmwllnsppmktcptlypvd 480
QY 481 NLPTKFTSYVRN 492
Db 481 nlptkftsyrn 492

RESULT 3
AAW37734
ID AAW37734 standard; Protein: 497 AA.
XX
AC AAW37734;
XX
DT 07-JUL-1998 (first entry)
XX
DE Human cytochrome P450RAI protein.
XX
KW Retinoid regulated gene; cytochrome P450 gene; enzyme;
XX
KW oxidative metabolism; P450RAI; retinoic acid; RA; promoter.
XX
OS Homo sapiens.
XX
PN MO9749832-A2.
XX
PD 31-DEC-1997.
XX
PF 23-JUN-1997; 97WO-CA00488.
XX
PR 01-OCT-1996; 96US-0724466.
XX
PR 21-JUN-1996; 96US-0667546.
XX
PA (TOOH ) UNIV QUEBENS KINGSTON.
XX
PI Petkovich PM;
XX
DR MPI; 1998-077193/07.
XX
DR N-PSDB; AAV09247.
XX
PT Identifying DNA encoding inducible or suppressible cytochrome P450 -
XX
PT by screening for drugs which reduce the catabolism of retinoic acid,
XX
PT useful in cancer chemotherapy and the treatment of acne and
XX
PT psoriasis
XX
PS Example 4; Pages 58-59; 113pp; English.
XX
CC This is the amino acid sequence of the human cytochrome P450RAI. Its
XX
CC expression is dependent on the presence of retinoic acid (RA). The
XX
CC retinoid-regulated genes such as the inducible cytochrome P450RAI
XX
CC gene specifically metabolises a derivative of the RA. The cytochrome
XX
CC P450 gene in general produces enzymes involved in the oxidative
XX
CC metabolism of endogenous and exogenous compounds. The cytochrome
XX
CC P450 nucleotide sequence can be used to induce or suppress the
XX
CC expression of its protein. P450RAI is highly induced by RA in cell
XX
CC lines and tissues. This allows for development of a drug screen
XX
CC using promoters and nucleotide sequences to identify drugs which are
XX
CC useful for reducing the catabolism of RA.
XX
SQ Sequence 497 AA;

```

Query Match 67.8%; Score 1738; DB 19; Length 497;
 Best Local Similarity 68.0%; Pred. No. 3,66-151;
 Matches 338; Conservative 63; Mismatches 84; Indels 12; Gaps 5;

```

QY 1 MGLTYLMTFLCTIVLPVLLFLAAKYLKEMLMIRVDPNCRSLPFGTGLPFGTGLQ 60
||| : : ||| ||||| : : : : : ||| ||| ||| ||| |||
Db 1 mgltylmtflctivlpvllflaalklwlvcvsgdrdscslpfpptmfpfifgletlm 60
QY 61 ILQRRKFLMKRQKYGYCTKTHLEGNPTVRMGADNVRQQLLGHKLVSQVPAVFTIL 120
||| : : : ||| : : ||| ||||| : : ||| ||||| : : ||| |||||
Db 61 vlgrckflgmkrkygfllykthlfgprlvrmgadnvrlllgddrlvsvhpasvrtll 120
QY 121 GSDTLNHNAGVQHKKKKRAIMAPSRDALNHYIVYIOEVKSAIOEMLQKDS- - -VIV 176
||| : : : ||| : : ||| ||||| : : ||| : : ||| : : |||
Db 121 gsgdlnhndsshkqkrkvlmatsrealceyvpvleeagssilegw- - -scqrgllv 177
QY 177 YPEMKLMFRIAMRIILGFEPEQIKTD- - -EQLVEAFEEEMIKLFLPIDVPSSGLYRG 233
||| : : ||||| : : ||| : : ||||| : : ||||| : : |||||
Db 178 ypevkrlmfriamrillgcep-qlagdgsdqglveafemtnrlfslpdyvpsglyrg 236
QY 234 LRAKNFIHSKIENIRKIKQD-DDNENQOKKDALQQLLIEHSRSDPFSIQAMKEATE 292
::||| : : : ||| : : ||| ||||| : : ||| : : ||| : : |||
Db 237 mkarnlharlnleqnrlaklqglraseagqckdalqlllehsrgerlmdqalkgsste 296
QY 293 LLEGGHETASTATSLAVMFLGITEVQKREVEQVEKEMQMTPGKGLSMELDLQKYT 352
||||| : : ||||| : : ||| : : ||||| : : ||| : : |||||
Db 297 llfgbettaasatsllyglpvhvlgkveelkskglcksnqddklmeilleqkyl 356
QY 353 GCYIKETLRINPVPVGGFRVALKTFELNGYQIPKGMNVIYSICDTHVDADVFPNKEEP 412
||||| : : ||||| : : ||||| : : ||||| : : ||||| : : |||||
Db 357 gcvikeetlrinppvpvgfrvalkftelngyqipkgnviysicdthvaelftkfeinp 416
QY 413 ERPFMSKLEDSGRFNYPFGGSRMCGVEKFAVLKIFLVELTQHNMILNSGPPMTK 472
||| : : ||| : : ||||| : : ||||| : : ||||| : : |||||
Db 417 drfsaphedastfslpfgglrscvgekafakllkifvelarhcdwqlngpplmkt 476
QY 473 GPITYPVNMLPTKFTSY 489
||| : : ||| : : ||| : : ||| : : ||| : : |||
Db 477 sptvypvdnlparfchf 493

```

```

RESULT 4
AAW44160
ID AAW44160 standard; Protein: 497 AA.
XX
AC AAW44160;
XX
DT 22-JUN-1998 (first entry)
XX
DE Human retinoid metabolising protein hP450RAI.
XX
KW Retinoid metabolising protein: P450RAI; retinoid oxidase;
XX
KW retinoic acid; human; inhibitor; antibody; cancer;
XX
KW actinic keratosis; oral leukoplakia; head tumour; neck tumour;
XX
KW non-small cell lung carcinoma; basal cell carcinoma;
XX
KW acute promyelocytic leukaemia; skin cancer; acne; psoriasis;
XX
KW ichthyosis; therapy; diagnosis; screening.
XX
OS Homo sapiens.
XX
PN MO9749915-A1.
XX
PD 31-DEC-1997.
XX
PF 23-JUN-1997; 97WO-CA00440.
XX
PR 01-OCT-1996; 96US-0724466.
XX
PR 21-JUN-1996; 96US-0667546.
XX
PA (TOOH ) UNIV QUEBENS KINGSTON.
XX
PI Beckett BR, Jones G, Petkovich PM, White JA;
XX
DR MPI; 1998-077178/07.
XX
DR N-PSDB; AAV12204.
XX
PT Retinoid metabolising protein - useful to develop products to treat,

```

PT e.g. cancer, actinic keratosis, oral leukoplakia, acne, psoriasis or
 XX ichthyosis

PS Claim 1; Page 54-55; 110pp; English.

XX This protein comprises a novel human retinoid metabolizing protein,
 CC designated hP450RAI. Its amino acid sequence was deduced from a
 CC cDNA clone (see AAV12204) isolated from a retinoid acid-treated NT2
 CC cell library. It includes a haem-binding motif characteristic of
 CC cytochrome P450 proteins. hP450RAI is a retinoid oxidase that has
 CC the ability to hydroxylate retinoic acid at the 4 position of the
 CC beta-ionone ring, and is inducible in epithelial cells exposed to
 CC retinoic acid. Zebrafish, human and mouse P450RAIs (see AAM4159-61)
 CC are claimed. They can be expressed in host cells and used to
 CC metabolize retinoic acid in an organism or cell, in drug screening,
 CC and to raise antibodies useful for inhibiting retinoic acid
 CC hydroxylation for the treatment of cancer, actinic keratosis, oral
 CC leukoplakia, secondary tumours of the head and/or neck, non-small
 CC cell lung carcinomas, basal cell carcinomas, acute promyelocytic
 CC leukaemia, skin cancer, and premalignancy associated with actinic
 CC keratosis, acne, psoriasis and/or ichthyosis.

XX Sequence 497 AA;

Query Match 67.8%; Score 1738; DB 19; Length 497;

Best Local Similarity 68.0%; Pred. No. 3,6e-151;

Matches 338; Conservative 63; Mismatches 84; Indels 12; Gaps 5;

QY 1 MGLYTLMTVEICTIVLPVLLFLAANKLWEMLMIRVDPNCRSPLPFGTGLPIGTGLQL 60
 DB 1 mglpallasaalcftvlpillflaalklwdlycvsgdrscalcplpypmgfpgfgetlqm 60
 QY 61 ILORRKLFRMRKRCYKCYTHLEGNPTVYMGADNVROILGHEHKLVSQVMPASVETIL 120
 DB 61 vlgtrkflgmkrkygfyfchlfgrptvymgadnvrrlllgdrlvswhmpasvrtll 120
 QY 121 GSDTLNVHGVQHKRKKALMRAFSDALEHYIPVIOEQVSAIQEWLQDSC---VLV 176
 DB 121 gsgclsnlhshskqkkykvmrafsrealecyvpteevsgslqwl---scgergliv 177
 QY 177 YPEMKRMFRIMRILGFEPEQIKTD---EDQLVAFEMIKNLSLPIVDFPSGLRYG 233
 DB 178 ypevkrmlfrimrlllgcep-qldagdgseqqlveafemtrnlslpdpvfgslryg 236
 QY 234 LRAARNFHSKIEENIRKKIQD--DDNENQKYYKDALQLLIENSRRSDEPFSLOAKKEATE 292
 DB 237 mkarnlharleqnirakloglrseagqckdaqlillehswgerldmqalqkssste 296
 QY 293 LLEGHETASTATSLVMPGLNTEVYVQKREVOEKVEMGMTPGKGLSMELLDOLKTT 352
 DB 297 lileghetastatsllvlglyphvlgvreekskyllicksnqdkldmetleqlkyi 356
 QY 353 GCYIKETLRINPEVPGFRVALKTEFELNGYQIPLKGMNVYSICDTHDVAVDPNKEEFP 412
 DB 357 gcyiketlrinpevpgfrvalktfelngyqplqkgnvyslctdhvaeiftnkefnp 416
 QY 413 ERMMSGLDEGSEFNIPFGGSRMCKVGEFAVLLKTLVELTQCNMILNSGPPTMT 472
 DB 417 drrmsgldegsefnipfggsgsmckvgefavllktlveltqcnmilsngpptmt 476
 QY 473 GPTIYVNDLPTKTSY 489
 DB 477 sptiyvndlptktsf 493

RESULT 5
 AAM37735
 ID AAM37735 standard; Protein; 497 AA.

AC AAM37735;
 XX
 DT 07-JUL-1998 (first entry)

XX Cytochrome P450RAI isoform.

XX Retinoid regulated gene; cytochrome P450 gene; enzyme;

KM oxidative metabolism; P450RAI; retinoic acid; RA; promoter; isoform.

XX Synthetic.

XX WO9749832-A2.

XX 31-DEC-1997.

XX 23-JUN-1997; 97WO-CA00488.

XX 01-OCT-1996; 96US-0724466.

XX 21-JUN-1996; 96US-0667546.

XX (TOOH) UNIV QUEBENS KINGSTON.

XX petkovich PM;

XX WPI; 1998-077193/07.

XX N-PSDB; AAV09252.

XX Identifying DNA encoding inducible or suppressible cytochrome P450 -

XX by screening for drugs which reduce the catabolism of retinoic acid,

XX useful in cancer chemotherapy and the treatment of acne and

XX psoriasis

XX Disclosure; Pages 596-598; 113pp; English.

XX This amino acid sequence is of an isoform of cytochrome zp450RAI.

XX Its expression is dependent on the presence of retinoic acid (RA).

XX The retinoid-regulated genes such as the inducible cytochrome P450RAI

XX gene specifically metabolises a derivative of the RA. The cytochrome

XX P450 gene in general produces enzymes involved in the oxidative

XX metabolism of endogenous and exogenous compounds. The cytochrome P450

XX nucleotide sequence can be used to induce or suppress the expression

XX of its protein. P450RAI is highly induced by RA in cell lines and

XX tissues. This allows for development of a drug screen using promoters

XX and nucleotide sequences to identify drugs which are useful for

XX reducing the catabolism of RA.

XX Sequence 497 AA;

Query Match 67.7%; Score 1734; DB 19; Length 497;

Best Local Similarity 67.7%; Pred. No. 8,4e-151;

Matches 336; Conservative 61; Mismatches 89; Indels 10; Gaps 4;

QY 1 MGLYTLMTVEICTIVLPVLLFLAANKLWEMLMIRVDPNCRSPLPFGTGLPIGTGLQL 60
 DB 1 mglpallasaalcftvlpillflaalklwdlycvsgdrscalcplpypmgfpgfgetlqm 60
 QY 61 ILORRKLFRMRKRCYKCYTHLEGNPTVYMGADNVROILGHEHKLVSQVMPASVETIL 120
 DB 61 vlgtrkflgmkrkygfyfchlfgrptvymgadnvrrlllgdrlvswhmpasvrtll 120
 QY 121 GSDTLNVHGVQHKRKKALMRAFSDALEHYIPVIOEQVSAIQEWLQDSC---VLV 176
 DB 121 gsgclsnlhshskqkkykvmrafsrealecyvpteevsgslqwl---scgergliv 177
 QY 177 YPEMKRMFRIMRILGFEPEQIKTD---EDQLVAFEMIKNLSLPIVDFPSGLRYG 234
 DB 178 ypevkrmlfrimrlllgcep-qldagdgseqqlveafemtrnlslpdpvfgslryg 237
 QY 235 LRAARNFHSKIEENIRKKIQD--DDNENQKYYKDALQLLIENSRRSDEPFSLOAKKEATEL 293
 DB 238 mkarnlharleqnirakloglrseagqckdaqlillehswgerldmqalqkssste 297
 QY 294 LLEGHETASTATSLVMPGLNTEVYVQKREVOEKVEMGMTPGKGLSMELLDOLKTT 353
 DB 298 lileghetastatsllvlglyphvlgvreekskyllicksnqdkldmetleqlkyi 357

PS Claim 1; Page 10; 13pp; Japanese.

CC The present sequence represents a human protein, designated PSEC64, which
 CC is related to neuron growth. The PSEC64 protein and its gene can be used
 CC for the development of a preventive agent for use in the treatment of
 CC diseases in which nerves are involved.

XX Sequence 216 AA:

Query Match 29.4%; Score 752.5; DB 21; Length 216;
 Best Local Similarity 68.2%; Pred. No. 3.9e-61;
 Matches 150; Conservative 28; Mismatches 31; Indels 11; Gaps 4;

QY 1 MGILYLMVETLTVLPVLLFLAAYKIMEMLMIRVDPNCRSPPLPGTMGLPFIETQLI 60
 DB 1 mgpallaalactcfvplllflaaklkwlycvsgdrscalplppgtngfffgtclgm 60
 QY 61 ILORRKFELMKROKGCICVTHLFGNPTVYMGADNVRQILGHEKIVSVQMPASVRTL 120
 DB 61 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 120
 QY 121 GSDTLNVHGVCHKNKKAIMRAFSDALEHYIPVIOEVSALQEWLQKSC---VLV 176
 DB 121 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 176
 QY 121 vlgtrfqlmqtkrkygflkthlfgpvtvmgadnvrlllgentlvsvmpasvrtll 120
 DB 121 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 120
 QY 177 YPEMKLMFRIMRILLGFEPEOIKTD---RQELVEAPEE 213
 DB 177 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 213
 QY 178 YPEMKLMFRIMRILLGFEPEOIKTD---RQELVEAPEE 216
 DB 178 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 216

RESULT 8

AAW27153
 ID AAW27153 standard; Protein: 472 AA.

XX AAW27153;

XX 14-APR-1998 (first entry)

XX Arabidopsis thaliana cytochrome P450-type hydroxylase.

XX Cytochrome P450-type hydroxylase; identification; brassinosteroid;
 KW brassinosteroid inhibitor; modified plant; recombinant production;
 KW testosterone.

XX Arabidopsis thaliana.

OS WO9735986-A1.

XX 02-OCT-1997.

XX 27-MAR-1997; 97WO-EP01586.

XX 27-MAR-1996; 96US-0622166.

XX (PLAC) MAX PLANCK GES. FOERDERUNG WISSENSCHAFTEN.

XX Altman T, Koncz C, Mathur J, Szekeres MA;

XX WPI: 1997-489649/45.

XX N-PSDB; AAT85306, AAT85307.

XX New isolated plant cytochrome P450-type hydroxylase gene - used to
 PT identify substances acting as brassino-steroid(s) or brassinosteroid
 PT inhibitors for the production of modified plants

XX Claim 1; Pages 44-46; 77pp; English.

XX The present sequence is Arabidopsis thaliana cytochrome
 CC P450-type hydroxylase. The hydroxylase can be used to identify
 CC brassinosteroids or brassinosteroid inhibitors, useful to produce
 CC plants with modified physiological and/or phenotypic
 CC characteristics. The modified plants may show, e.g. stimulation of

CC growth, increased cell elongation, increased wood production,
 CC accelerated seed germination at low temperatures, an increase in
 CC dry weight, repressed anthocyanin production during growth in light
 CC and/or inhibited de-etiolation which is induced, e.g. by cytokinin,
 CC in the dark or an increase in stress tolerance. The hydroxylase or
 CC its coding sequence can also be used for the recombinant production
 CC of compounds, e.g. testosterone.

XX Sequence 472 AA:

Query Match 21.6%; Score 552.5; DB 18; Length 472;
 Best Local Similarity 29.9%; Pred. No. 3e-42;
 Matches 145; Conservative 98; Mismatches 205; Indels 37; Gaps 14;

QY 19 LFLFAAVKIMEMLMIRVDPNCRSPPLPGTMGLPFIETQLI-----LORRKFELMKRO 73
 DB 7 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 73
 QY 74 KYGCITVTHLFGNPTVYMGADNVRQILGHEKIVSVQMPASVRTIIGSDTLNVHGVCH 133
 DB 74 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 133
 QY 66 rygsvfmthlfgpvtvmgadnvrlllgentlvsvmpasvrtlllgentlvsvmpasvrtll 125
 DB 66 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 125
 QY 134 KKKKKAIMRAFSDAL--EHYIPVIOEVSALQEWLQKSCVLYPEMKLMFRIMARI 191
 DB 134 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 191
 QY 126 k-rmbsltmsfamsislkdhlmldidrlvrfnldsw---srrvlmeaakktfcltqk 181
 DB 126 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 181
 QY 192 LIGFEP-----EQITDQELVEAFEMIKNLFSLPIDVPSGLYR-GLRARNFTHSKIEE 246
 DB 192 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 246
 QY 182 lmsldpgweslkeyllvlegf-----fslpjp1-ftltrykaidarrfrvaaliv 233
 DB 182 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 233
 QY 247 NIKKIDDDDNENQKKDALQLLITENSRRSDPEFSLOAKKEATELLFGHETASTAT 306
 DB 247 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 306
 QY 234 vvmkr-reeeesgerkkmaall-----aaddfseelvdflvalvagyttctimt 288
 DB 234 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 288
 QY 307 SLVAFGLNTEVOKVEBEVOKVEMGMTPGKLSKELLQKTYGVKTELRINPPY 366
 DB 307 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 366
 QY 289 lavkflteplalaglkke-hekir-amksdyslswdykmpftqcvneclrvana11 346
 DB 289 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 346
 QY 367 PGGRVALKPELNGYOIPKGMNYISICDTHVADVFPKKEEFOPEREKSLIEDGSRP 426
 DB 367 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 426
 QY 347 ggvtrrmtdelkyklpkykvsfrravhldpnhfkarfnprwsgnsvtgtgpn 406
 DB 347 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 406
 QY 427 NYTFPGGSRMCVGEKAFKYLKFLVELTHOCHWILSNCPPTMKGPTIYPVDNLPTAF 486
 DB 427 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 486
 QY 407 vtfpfggprlcpyselaravalsvflhrlytfgiswv-----paegdklyfptrtqky 461
 DB 407 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 461
 QY 487 TSYVR 491
 DB 487 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 491
 QY 462 plfvk 466
 DB 462 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 466

RESULT 9

AAG44571
 ID AAG44571 standard; Protein: 472 AA.

XX AAG44571;

XX 18-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 55847.

XX Protein identification; signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KW termination sequence.

XX Arabidopsis thaliana.

OS EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132853.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139453.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140635.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.

PR 12-JUL-1999; 99US-0142927.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 19-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 23-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145919.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 15-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.

OY	134	KNKKKAIIRAPASRAL--EHYIPVIOEVSASIAOEWLQKSCVLYVPEKKKIMFRIAMRI	191
Db	126	k-rnbslmsfamsnsilkdhlmldidlvrlfnldsw---srrylimeakkifefaltvqk	181
OY	192	LLGFEF---EQIKTDQDELVEAFEEEMIKNLFSLPIDVPEFSGLYR-GLRAINFHSKIEE	246
Db	182	lmsldpgeawsesslkkeyllvylegf-----fsbjpll-fstlyrkaldgrfrvaaltv	233
OY	247	NIRKKIOODDNNENOKKADLQOLIENSRRSDPEFSLOAKKEAATELLPGSHETASTAT	306
Db	234	vmkrr-reeeeggerkkkmhaall-----aaddgfdeeeivdvlvialvaygetcsint	288
OY	307	SLVMEFLGINTEVOKRVEEVOEKVEKMGMYPPGKGLSHELLDQLKYGVCIKETLRINPV	366
Db	289	lavflfletptialaglkkee-hekir-amsksdyslswsdykmpfcgvvnetlrvanli	346
OY	367	PGGFRVAKTTEFLGNGYOIPKGMWNTYISICDTHVADVPKKEEFOERFMSKLEDSRF	426
Db	347	ggvfrtrantdvclkykklpkykswkfsfravhldpnlkardatfrnpwsgnsvtgpsn	406
OY	427	NYIPFGGSGRMVCVGEKFAKYLKIFVELTHQCNWILSNPGPTMKTGPTTYPVDNLPTKE	486
Db	407	vftpfgggprlpcpygelaralavslfnrlvtgfsww-----paegdklvfptrtqky	461
OY	487	TSYVR	491
Db	462	plfvk	466
RESULT	11		
AAAG45021			
AAAG45021		standard; Protein; 491 AA.	
XX	AAAG45021:		
AC	AAAG45021:		
XX	18-OCT-2000	(first entry)	
XX	Arabidopsis thaliana	protein fragment SEQ ID NO: 56468.	
XX	Protein identification; signal transduction pathway; metabolic pathway;		
KW	hybridisation assay; genetic mapping; gene expression control; promoter;		
KW	termination sequence.		
XX	Arabidopsis thaliana.		
OS	EPI033405-A2.		
PM	06-SEP-2000.		
PD	25-FEB-2000; 2000EP-0301439.		
XX	25-FEB-1999;	99US-0121825;	
XX	05-MAR-1999;	99US-0123180.	
XX	09-MAR-1999;	99US-0123548.	
XX	23-MAR-1999;	99US-0125788.	
XX	25-MAR-1999;	99US-0126264.	
XX	29-MAR-1999;	99US-0126785.	
XX	01-APR-1999;	99US-0127462.	
XX	06-APR-1999;	99US-0128234.	
XX	18-APR-1999;	99US-0128714.	
XX	16-APR-1999;	99US-0129685.	
XX	19-APR-1999;	99US-0130077.	
XX	21-APR-1999;	99US-0130449.	
XX	23-APR-1999;	99US-0130510.	
XX	28-APR-1999;	99US-0130891.	
XX	30-APR-1999;	99US-0131449.	
XX	30-APR-1999;	99US-0132048.	
XX	04-MAY-1999;	99US-0132484.	
XX	05-MAY-1999;	99US-0132485.	
XX	06-MAY-1999;	99US-0132486.	
XX	06-MAY-1999;	99US-0132487.	
XX	07-MAY-1999;	99US-0132863.	
PR	11-MAY-1999;	99US-0134256.	
PR	14-MAY-1999;	99US-0134218.	
PR	14-MAY-1999;	99US-0134219.	
PR	14-MAY-1999;	99US-0134221.	
PR	14-MAY-1999;	99US-0134370.	
PR	18-MAY-1999;	99US-0134768.	
PR	18-MAY-1999;	99US-0134941.	
PR	20-MAY-1999;	99US-0135124.	
PR	21-MAY-1999;	99US-0135353.	
PR	24-MAY-1999;	99US-0135629.	
PR	25-MAY-1999;	99US-0136021.	
PR	27-MAY-1999;	99US-0136392.	
PR	28-MAY-1999;	99US-0136782.	
PR	01-JUN-1999;	99US-0137822.	
PR	03-JUN-1999;	99US-0137528.	
PR	04-JUN-1999;	99US-0137502.	
PR	07-JUN-1999;	99US-0137724.	
PR	08-JUN-1999;	99US-0138094.	
PR	10-JUN-1999;	99US-0138840.	
PR	10-JUN-1999;	99US-0138647.	
PR	14-JUN-1999;	99US-0139119.	
PR	16-JUN-1999;	99US-0139452.	
PR	16-JUN-1999;	99US-0139453.	
PR	17-JUN-1999;	99US-0139492.	
PR	18-JUN-1999;	99US-0139454.	
PR	18-JUN-1999;	99US-0139455.	
PR	18-JUN-1999;	99US-0139456.	
PR	18-JUN-1999;	99US-0139457.	
PR	18-JUN-1999;	99US-0139458.	
PR	18-JUN-1999;	99US-0139459.	
PR	18-JUN-1999;	99US-0139460.	
PR	18-JUN-1999;	99US-0139461.	
PR	18-JUN-1999;	99US-0139462.	
PR	18-JUN-1999;	99US-0139463.	
PR	18-JUN-1999;	99US-0139750.	
PR	18-JUN-1999;	99US-0139763.	
PR	21-JUN-1999;	99US-0139617.	
PR	22-JUN-1999;	99US-0139699.	
PR	23-JUN-1999;	99US-0140353.	
PR	23-JUN-1999;	99US-0140354.	
PR	24-JUN-1999;	99US-0140695.	
PR	24-JUN-1999;	99US-0140823.	
PR	26-JUN-1999;	99US-01409	

```

PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145291.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145919.
PR 02-AUG-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147132.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159325.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.

```

```

PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161932.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

Query Match 21.6%; Score 552.5; DB 21; Length 491;
 Best Local Similarity 29.9%; Pred. No. 3,2e-42;
 Matches 145; Conservative 98; Mismatches 205; Indels 37; Gaps 14;

```

QY 19 LLEFLAVALKLEMLIRYVDPCRSFPLPGTGLPPIGTTLOL-----LORRFLRRKKQ 73
DB 26 LLLSSlaagfllllrrrrr-rmgplpsslgpllgelqlyayktenpepflderiva 84
QY 74 KYGCIYKTHLFGNPTVRWAGADNVRQILGSEKLVSVQMPASVFTIIGSDPLSVHGVQH 133
DB 65 rygsyvmchlfgeptifcaadpetnrfvlyqnegklfecypasicnllgkhslllmkyslh 144
QY 134 KKKRAIMRAESRAL--EHYIPVIOGEVKSALIOEWLOKDSCVLVPEMKKLMFRIMRI 191
DB 145 k-rmhsllmsfansllkdhlmldldtrfndsw--ssrvllmeaakkltfelvkg 200
QY 192 LGFEP---EQIKTDEDELVEAFEMIKNLSPIVDPFSGLYR-GLRARNFHskIEE 246
DB 201 lmsdpgeysesllrkeyllvlegf-----fsllppl-istcykaigaarkaaaltv 252
QY 247 NIKKKIODDNEBQKYDALQILTENSRSDEPSSLOAMKEATELLFGSHETASTAF 306
DB 253 vvmkr-reeeegaerkkdmllaill---aaddgfsdeelvdfvallvaygetstlmt 307
QY 307 SLVMEFLGNTVEYVQKVRREVOEKVEMGMVTPGKGLSMELDLQLYTGCYVKEPTLRINPV 366
DB 308 lsvkflteplalaqikee-hekir-amksdsyslswdyksmpfctgvvnellyvaall 365
QY 367 PGCFYALKTEPLNGYQIPKGMNVITYSIDPHYADVFPNKEEPQRPFRMSKGLIEDSRF 426
DB 366 ggyframdvleikgklypkqkwkfsstravhldpnhktartfnpwqsnsvtqpsn 425
QY 427 NYIPFGGSRMGVGEKFAKVALIKIRFVLELTHQCNWILTSNGPPTMKGTPTIYPVDNLPTK 486
DB 426 vtfptggyprlcpgyelarvalsvflhrvtgfsvw----paeqdklvifprrtqky 480
QY 487 TSYVR 491
DB 481 plfvk 485

```

RESULT 12
 AAG44570
 ID AAG44570 standard; Protein: 492 AA.
 XX
 AC AAG44570;
 XX
 DT 18-OCT-2000 (first entry)
 XX
 XX Arabidopsis thaliana protein fragment SEQ ID NO: 55846.
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 55846.
 XX
 XX Protein identification; signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KW termination sequence.
 XX
 OS Arabidopsis thaliana.
 XX
 XX EP1033405-A2.
 XX
 XX
 PD 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.
PF XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126785.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129645.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139482.
PR 18-JUN-1999; 99US-0139434.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.

PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146389.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 05-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149920.
PR 25-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.


```

PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 14-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

Query Match 21.6%; Score 552.5; DB 21; Length 492;

Best Local Similarity 29.9%; Pred. No. 3, 2e-42; Mismatches 205; Indels 37; Gaps 14;

```

Matches 145; Conservative 98; Mismatches 205; Indels 37; Gaps 14;

QY 19 LLELAAYKLEMLIRVDPNCRSPPLPRTGMLPFGTETQLI-----LQRRKFLRRKQ 73
DB 27 LLLLSIAAGFIILLIRTRY-rmgllpbgsglplgetfglqayktenpeptideva 85
QY 74 KGCGLYKTHLGNPVYRMGADNVRQILLGHEKLVSQMPASVFTIIGSDTLMSVHGVOH 133
DB 86 RYGSYFMCHLIGEPCLIFSADEPCNRFVLQNEGKLFECSYPASICNLIQKHSIILMKGSLH 145
QY 134 KKKRAIMKAFSRDAL--EHYIPVIOEFVKSALQEWLQKDCVLYEPMKKLMFRIMRI 191
DB 146 k-rmlslfmsfnsalldkhlmlddrldrftldsw---srrvllmeaakkltfelvq 201
QY 192 LIGEP-----EQIKRDEGLVEAFEMIKNLFSLPIVDFPSGLYR-GLRARNFHSKIEE 246
DB 202 lmsltgswesellrkeylllvlegf-----fsllplrl-fstkykaidarkwaaalv 253
QY 247 NIKRKIODDNEKEQYDALQLLIENSRRSDPEFSIQAKMEATTELFEGHETTASTAT 306
DB 254 vvmkr-reeeegaeerkdkmaiall---aaddgfsdeidvdlvallygetstclmt 308
QY 307 SLVYMLGLNTEVQVREVOEKVEMGNTTPGKGLSMELDLQLYTTCVYKELIRINPV 366
DB 309 lsvkllletplajqlkee-heklir-amksdsyslswdyksmpftgcynveltrvanil 366
QY 367 PCGFVALKTEFLNGYQIPKGMNVLYSICDTHVDADVFNKEEFPQERFMSKGLDEDSRF 426
DB 367 ggvframdvleikgkylpkqkvfssfravhlqpnhkdaarltlnprwqsnstltgpn 426
QY 427 NIIPIFGGSRKVCVKEFKVLKITIVELTGHCHWILISNGPPTMKTGFTIYVDNLPTKF 486
DB 427 vltptfggprlcpqygelarvalsvflhrllvlgfsw-----paegdklvfprltltkry 481

```

```

QY 487 TSYVR 491
DB 482 plfvk 486

```

RESULT 13

AA644572 standard; Protein; 444 AA.

AC AA644572;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 55848.

KW Protein identification; signal transduction pathway; metabolic pathway; hybridisation assay; genetic mapping; gene expression control; promoter; termination sequence.

OS Arabidopsis thaliana.

PN EP1033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

```

PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0138847.
PR 16-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.

```

[illegible]

Db 63 fvlqneqkLfecsyPasicnllgkhsjllmkgsJhk-rmbshlmsfnsislikhlmldi 121
QY 157 QOEVSASIAQELQKQDSCLVLPPEKKLMFRAMILGFEF-----EQIKDEQELVEAF 212
Db 122 dlvfnldsw---ssrvllmeakkltfclvqqlmsfdpgwseelrkeyllvlegf- 177
QY 213 EMKKLFLSLPDPVPSGLYR-GLRARNFINSKLEENIRKKIQQDDNENKRYDALQLLI 271
Db 178 -----fslppl-fstlyrkaigarkvaeallvvymkr-reeeegaarckmdlaal 229
QY 272 ENSRSDPEPSIQAMKEATLLEFGHETASTSTSLVMEGLNTEVYQVREVEQKVE 331
Db 230 ---aaddgfdeedvldvalvaygetstlmrlavkltetplalaqlkee-hekir 284
QY 332 MGMTPGKGLSMELIDQKTYGCVIKETLRINPPVPGCFRVALTKTFELNGYQIIPKGMNVI 391
Db 285 -amkdsyslswsdykmsrpfqcvneclrvanllgsvframdvclkykklpkqkvf 343
QY 392 YSICTHVADVPPNKEFOPPERMSKGLDEGRFNITIPFGGSRMGCVGEFAKVLIKIF 451
Db 344 sflravhldpnhfkdarclfnprwqsnsvltgpnvltptgysgrllopyelarvalsvf 403
QY 452 LEVELTOHNMILSNCPMTKPTIYPVDNLPTKFTSYVR 491
Db 404 lhrvtgfsw----paegdklvffpotttrqkrpflvk 438

RESULT 14

AAG45023 standard; Protein: 444 AA.

AAG45023;

18-OCT-2000 (first entry)

Arabidopsis thaliana protein fragment SEQ ID NO: 56470.

Protein identification; signal transduction pathway; metabolic pathway;
hybridisation assay; genetic mapping; gene expression control; promoter;
termination sequence.

Arabidopsis thaliana.

EP1033405-A2.

06-SEP-2000.

25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139432.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0144864.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2001, 10:41:47 ; Search time 88.14 Seconds
(without alignments)

125.614 Million cell updates/sec

Title: US-09-668-482-2

Perfect score: 2563
Sequence: 1 MGLYTLMVFLCTIVLPVLL.....GPTIYPVDNLPTKFTSVRN 492

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
Maximum Match 1008
Listing first 45 summaries

Database : Issued Patents_AA:*

1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCRTUS.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2563	100.0	492	3	US-08-724-466B-2
2	2563	100.0	492	4	US-08-882-164D-2
3	1738	67.8	497	3	US-08-724-466B-4
4	1738	67.8	497	4	US-08-882-164D-4
5	1734	67.7	497	4	US-08-882-164D-32
6	552.5	21.6	472	2	US-08-822-166A-2
7	552.5	21.6	472	2	US-08-622-166A-4
8	343	13.4	511	4	US-08-991-677-4
9	339.5	13.2	490	1	US-08-201-118-3
10	339.5	13.2	490	1	US-08-201-118-9
11	339.5	13.2	490	2	US-08-238-821B-3
12	339.5	13.2	490	2	US-08-238-821B-9
13	339.5	13.2	490	5	PCT-US95-05744-3
14	339.5	13.2	490	5	PCT-US95-05744-9
15	330.5	12.9	520	2	US-09-091-432-2
16	322.5	12.6	500	2	US-08-314-601-2
17	322.5	12.6	500	5	PCT-US95-13051-2
18	316.5	12.3	490	1	US-08-101-118-13
19	316.5	12.3	490	2	US-08-238-821B-13
20	316.5	12.3	490	5	PCT-US95-05744-13
21	311.5	12.2	490	1	US-08-201-118-1
22	311.5	12.2	490	2	US-08-238-821B-1
23	311.5	12.2	490	5	PCT-US95-05744-1
24	302	11.8	510	3	US-08-948-564-4
25	296	11.5	496	1	US-08-313-075A-50
26	293.5	11.5	490	1	US-08-201-118-7
27	293.5	11.5	490	2	US-08-238-821B-7

28	293.5	11.5	490	5	PCT-US95-05744-7	Sequence 7, Appl1
29	287	11.2	504	1	US-08-457-274A-25	Sequence 25, Appl1
30	287	11.2	504	5	PCT-US95-05758-25	Sequence 25, Appl1
31	282.5	11.0	476	1	US-08-313-075A-30	Sequence 30, Appl1
32	275.5	10.7	500	4	US-09-292-768-68	Sequence 68, Appl1
33	274.5	10.7	500	4	US-09-292-768-4	Sequence 4, Appl1
34	271.5	10.6	500	4	US-08-881-784-9	Sequence 9, Appl1
35	271.5	10.6	500	4	US-09-292-768-70	Sequence 70, Appl1
36	268.5	10.5	490	1	US-08-201-118-11	Sequence 11, Appl1
37	268.5	10.5	490	2	US-08-238-821B-11	Sequence 11, Appl1
38	268.5	10.5	490	5	PCT-US95-05744-11	Sequence 11, Appl1
39	268.5	10.5	516	3	US-08-948-564-12	Sequence 12, Appl1
40	268.5	10.5	521	3	US-08-948-564-14	Sequence 14, Appl1
41	267.5	10.4	490	1	US-08-201-118-5	Sequence 5, Appl1
42	267.5	10.4	490	2	US-08-238-821B-5	Sequence 5, Appl1
43	267.5	10.4	496	2	PCT-US95-05744-5	Sequence 5, Appl1
44	267.5	10.4	496	4	US-09-292-768-64	Sequence 64, Appl1
45	266	10.4	513	3	US-08-948-564-6	Sequence 6, Appl1

ALIGNMENTS

```
RESULT 1
US-08-724-466B-2
; Sequence 2, Application US/08724466B
; Patent No. 6063606
;
; GENERAL INFORMATION:
; APPLICANT: Petkovich, P. Martin, White, Jay A.,
; APPLICANT: Beckett, Barbara R., Jones, Glenville
; TITLE OF INVENTION: Retinoid Metabolizing Protein
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESS: Blake, Cassels & Graydon
; STREET: Box 25, Commerce Court West
; CITY: Toronto
; ZIP: M5L 1A9
; COUNTRY: Canada
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 Mb storage
; COMPUTER: COMPAQ, IBM PC compatible
; OPERATING SYSTEM: MS-DOS 5.1
; SOFTWARE: WORD PERFECT
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08724,466B
; FILING DATE: October 1, 1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/667,546
; FILING DATE: June 21, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Hunt, John C.
; REGISTRATION NUMBER: 36,424
; REFERENCE/DOCKET NUMBER: 50767/00004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 863-4344
; TELEFAX: (416) 863-2653
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 492 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
; US-08-724-466B-2
```

Query Match 100.0%; Score 2563; DB 3; Length 492;
Best local Similarity 100.0%; Pred. No. 1.8e-246;
Matches 492; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGLYTLMVFLCTIVLPVLLAFAVYKEMIMRRVDPNCRSPDPGTMGLPFGTTLQ 60
DB 1 MGLYTLMVFLCTIVLPVLLAFAVYKEMIMRRVDPNCRSPDPGTMGLPFGTTLQ 60

```

QY 61 ILORRFLRRKROKGYCICYKTHLFGNPTVYVMGADVNRQILLGEHKLVSQVMPASVYRTIL 120
DB 61 ILORRFLRRKROKGYCICYKTHLFGNPTVYVMGADVNRQILLGEHKLVSQVMPASVYRTIL 120
QY 121 GSDTSLNVHGVQHKNNKKAIIMRAFSDALEHYIPVIOEYKSAIOEWLQKDSCLVYYPEM 180
DB 121 GSDTSLNVHGVQHKNNKKAIIMRAFSDALEHYIPVIOEYKSAIOEWLQKDSCLVYYPEM 180
QY 181 KKLMEFIAMRILIGFEPEDIKTDEQELVEAFEEIMKNLFSLPIDVPPSGLYRGLRARNFT 240
DB 181 KKLMEFIAMRILIGFEPEDIKTDEQELVEAFEEIMKNLFSLPIDVPPSGLYRGLRARNFT 240
QY 241 HSKIEENIRKKIQQDDNENEOKYKDALQLLIENSRSDPEFSLOAKKEATELLFGGHET 300
DB 241 HSKIEENIRKKIQQDDNENEOKYKDALQLLIENSRSDPEFSLOAKKEATELLFGGHET 300
QY 301 TASTATSLVMFLGLNTEVQVKEEVOEKVEEMGMTPGKGLSMELLDOUKYTGCVIKETL 360
DB 301 TASTATSLVMFLGLNTEVQVKEEVOEKVEEMGMTPGKGLSMELLDOUKYTGCVIKETL 360
QY 361 RINPVPGGFRVALKTFEELNGYOIPKGMNVIYSICDTHVADVFPKKEFOBEREFSKGL 420
DB 361 RINPVPGGFRVALKTFEELNGYOIPKGMNVIYSICDTHVADVFPKKEFOBEREFSKGL 420
QY 421 EDGSRNRYIPFGGSGRMCGKKEFAVLLKIFLVELTQHNCNWLISNPPMTKGTPTIYPVD 480
DB 421 EDGSRNRYIPFGGSGRMCGKKEFAVLLKIFLVELTQHNCNWLISNPPMTKGTPTIYPVD 480
QY 481 NLPTKFTSYVRN 492
DB 481 NLPTKFTSYVRN 492

```

```

; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-882-164d-2
Query Match 100.0%; Score 2563; DB 4; Length 492;
Best local Similarity 100.0%; Pred. No.1.8e-246;
Matches 492; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGLYTLMTVFLCTIVLPVLLFLAAYKLMEMLMIRVDPNCRSPLPPTGTMGLPFIGETIQL 60
DB 1 MGLYTLMTVFLCTIVLPVLLFLAAYKLMEMLMIRVDPNCRSPLPPTGTMGLPFIGETIQL 60
QY 61 ILORRFLRRKROKGYCICYKTHLFGNPTVYVMGADVNRQILLGEHKLVSQVMPASVYRTIL 120
DB 61 ILORRFLRRKROKGYCICYKTHLFGNPTVYVMGADVNRQILLGEHKLVSQVMPASVYRTIL 120
QY 121 GSDTSLNVHGVQHKNNKKAIIMRAFSDALEHYIPVIOEYKSAIOEWLQKDSCLVYYPEM 180
DB 121 GSDTSLNVHGVQHKNNKKAIIMRAFSDALEHYIPVIOEYKSAIOEWLQKDSCLVYYPEM 180
QY 181 KKLMEFIAMRILIGFEPEDIKTDEQELVEAFEEIMKNLFSLPIDVPPSGLYRGLRARNFT 240
DB 181 KKLMEFIAMRILIGFEPEDIKTDEQELVEAFEEIMKNLFSLPIDVPPSGLYRGLRARNFT 240
QY 241 HSKIEENIRKKIQQDDNENEOKYKDALQLLIENSRSDPEFSLOAKKEATELLFGGHET 300
DB 241 HSKIEENIRKKIQQDDNENEOKYKDALQLLIENSRSDPEFSLOAKKEATELLFGGHET 300
QY 301 TASTATSLVMFLGLNTEVQVKEEVOEKVEEMGMTPGKGLSMELLDOUKYTGCVIKETL 360
DB 301 TASTATSLVMFLGLNTEVQVKEEVOEKVEEMGMTPGKGLSMELLDOUKYTGCVIKETL 360
QY 361 RINPVPGGFRVALKTFEELNGYOIPKGMNVIYSICDTHVADVFPKKEFOBEREFSKGL 420
DB 361 RINPVPGGFRVALKTFEELNGYOIPKGMNVIYSICDTHVADVFPKKEFOBEREFSKGL 420
QY 421 EDGSRNRYIPFGGSGRMCGKKEFAVLLKIFLVELTQHNCNWLISNPPMTKGTPTIYPVD 480
DB 421 EDGSRNRYIPFGGSGRMCGKKEFAVLLKIFLVELTQHNCNWLISNPPMTKGTPTIYPVD 480
QY 481 NLPTKFTSYVRN 492
DB 481 NLPTKFTSYVRN 492

RESULT 3
US-08-724-466B-4
; Sequence 4, Application US/08724466B
; Patent No. 6063606
; GENERAL INFORMATION:
; APPLICANT: Petkovich, P. Martin, White, Jay A.,
; APPLICANT: Beckett, Barbara R., Jones, Glenville
; TITLE OF INVENTION: Retinoid Metabolizing Protein
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESS: Blake, Cassels & Graydon
; STREET: Box 25, Commerce Court West
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5L 1A9
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 Mb storage
; OPERATING SYSTEM: MS-DOS 5.1
; SOFTWARE: WORD PERFECT
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/724,466B
; FILING DATE: June 25, 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/667,546
; FILING DATE: June 21, 1996
; APPLICATION NUMBER: 08/724,466
; FILING DATE: October 1, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Hunt, John C.
; REGISTRATION NUMBER: 36,424
; REFERENCE/DOCKET NUMBER: 50767/00010
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 863-4344
; TELEFAX: (416) 863-2653
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 492 amino acids
; TYPE: amino acid
;
;

```



```

?      NAME: Hunt, John C.
?      REGISTRATION NUMBER: 36,424
?      REFERENCE/DOCKET NUMBER: 50767/00004
?      TELECOMMUNICATION INFORMATION:
?      TELEPHONE: (416) 863-4344
?      TELEFAX: (416) 863-2653
?      INFORMATION FOR SEQ ID NO: 4:
?      SEQUENCE CHARACTERISTICS:
?      LENGTH: 497 amino acids
?      TYPE: amino acid
?      STRANDEDNESS: single
?      TOPOLOGY: linear
?      US-08-724-466B-4

Query Match          67.8%; Score 1738; DB 3; Length 497;
Best Local Similarity 68.0%; Pred. No. 2,2e-164;
Matches 338; Conservative 63; Mismatches 84; Indels 12; Gaps 5

QY      1 MGIYTMVFLCTIVLIVLVLLTAAYKLMEMIRRVDPNCRSLPPLPGTMDLPIGTGLQ 60
        |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db      1 MGPIALLASALCTFVPLPLFLFLAIAIKLMDLVCVSGRDSRSCALPLPPTMGPPFGEGLQ 60

QY      61 ILQRKRLRKRCQVCICITKTHLPGNPIYRVYGADNVRQIILGEEKLYSVQMPASVETIL 120
        ::|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db      61 VLQRKRLKQKKRKRYGFIYKTHLPGRPIYRVYGADNVRRIILGDRILYSVMWPAVETIL 120

QY      121 GSDPLSNVHGVOHKRKKKAIMRAFSRDALEHYIPVIOGEVKSATIQEWLQKDS---VLV 176
        |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db      121 GSGCLSLNHDSSHQKKRYIMRAFSREALECYVPVITEGVSSLEQWL---SCGERGLV 177

QY      177 YPEKKIKMFIAKRIILGPEPQIKTD---RQELVAFEEEMIKNLSPIDVPSSGLYRG 233
        |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db      178 YPEKRIIMFIAKRIILGCEP-QLAGDGSQQLVEAEEMIRNLSPLDIVPSSGLYRG 236

QY      234 LRANFVHSKIKENIRKKIOD--DDNENEQKYDALQILIEHSRSDPEFSSLOAKKEATE 292
        ::|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db      237 MKAANLIHAIEQNIIRAKIQLGRASVAGGCKDALQILIEHSWEMGERLDMQALKOSTE 296

QY      293 LLEGGHTASTATSLMVLGLNTEYVQVREVEQKYEMGMYTPGKLSMELLDDQIKYT 352
        |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db      297 LLEFGHEHTASATSLITTYLGLYPRVLQVREBELKSKGLCKSNODNKLDMELLQOLKIYI 356

QY      353 GCYIKETLRINRPVPGGFRVALKTFEELNGYQIPKGNVYISICDHDVADVPKKEEFP 412
        |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db      357 GCYIKETLRINRPVPGGFRVALKTFEELNGYQIPKGNVYISICDHDVADLEFTKKEEFP 416

QY      413 ERFNSKGLDEGSRFNYIPFGGSRMCVGRKFAKVLKILFVELTHQCMWILSNBPWTKT 472
        ::|:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db      417 DRFSAPHEDEASRSRFPFGGSLRCVGRKEFAKILKIFVELARHCDWOLLNGPPTWKT 476

QY      473 GPITYPDNLPTKTSY 489
        |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db      477 SPYVYPVDNLPRPPTHF 493

RESULT      4
US-08-882-164D-4
; Sequence 4, Application US/08882164D
; Patent No. 630624
; GENERAL INFORMATION:
; APPLICANT: Petkovich, P. Martin, White, Jay A.,
; APPLICANT: Beckett, Barbara R., Jones, Glenville
; TITLE OF INVENTION: Retinoid Metabolizing Protein
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Blake, Cassels & Graydon
; STREET: Box 25, Commerce Court West
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5L 1A9
; COMPUTER READABLE FORM:

```

MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 Mb storage
 COMPUTER: COMPAQ, IBM PC compatible
 OPERATING SYSTEM: MS-DOS 5.1
 SOFTWARE: WORD PERFECT
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/882.164D
 FILING DATE: June 25, 1997
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/667.546
 FILING DATE: June 21, 1996
 APPLICATION NUMBER: 08/724.466
 FILING DATE: October 1, 1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Hunt, John C.
 REGISTRATION NUMBER: 36,424
 REFERENCE/DOCKET NUMBER: 50767/00010
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (416) 863-4344
 TELEFAX: (416) 863-2653
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 497 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear

[illegible]

```

GENERAL INFORMATION:
APPLICANT: Petkovich, P. Martin, White, Jay A.,
APPLICANT: Beckett, Barbara R., Jones, Glenville
TITLE OF INVENTION: Retinoid Metabolizing Protein
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESS: Blake, Cassels & Graydon
STREET: Box 25, Commerce Court West
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5L 1A9
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 Mb storage
COMPUTER: COMPAQ, IBM PC compatible
OPERATING SYSTEM: MS-DOS 5.1
SOFTWARE: WORD PERFECT
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/882,164D
FILING DATE: June 25, 1997
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 08/667,546
FILING DATE: June 21, 1996
APPLICATION NUMBER: 08/724,466
FILING DATE: October 1, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Hunt, John C.
REGISTRATION NUMBER: 36,424
REFERENCE/DOCKET NUMBER: 50767/00010
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 863-4344
TELEFAX: (416) 863-2653
INFORMATION FOR SEQ ID NO: 32:
SEQUENCE CHARACTERISTICS:
LENGTH: 497 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-882-164D-32

Query Match
Best Local Similarity 67.7%; Score 1734; DB 4; Length 497;
Matches 336; Conservative 61; Mismatches 89; Indels 10; Gaps 4;

QY 1 MGLYLTMTVETCTVLPVLLFLAAYKLMEMLRVDPNCRSPPLPGTGLPIGETIOL 60
DB 1 MGLPMLASALCTFVPLPLFLAALKMDLYCVSSRDSICALPLPGTGMGPFPGSTIOM 60
QY 61 ILQRRKFLMKRQKGCYIKTTLFGNPTVRVAGADVNRQIILGEHKLVSQMPASVRTIL 120
DB 61 VLQRRKFLQMKRRTGFIYKTHLFGFPTVRVAGADVNRILLEGHKLVSVMHPASVRTIL 120
QY 121 GSDTLNVHGVQHKKKRAIMRAFSDALEHYIPVIOEVKSAIOEMLQKDS---VLV 176
DB 121 GAGCLSNLHDSHQRKQKVTMQAFSREALQCVLVIAEVSVCLEML---SCGEGILV 177
QY 177 YPEMKKIFRIAMRILLGPEPQI--KTDEQLVEAFEMIKNLFSPLTIDVPFSGLYRGL 234
DB 178 YPEVRLMEFRIAMRILLGCEPDPAGGDEQOLVEAFEMTNTLSPLIDVPFSGLYRGL 237
QY 235 RARNIHSHKIEENIRKTIOD--DDNENQKXKDALQTLTENSRSDEPFLQMKKEATL 293
DB 235 RARNIHRIEENIRAKIRRLQATPEPDGCKDALQTLTENSWEGERLDMQALKOSTEL 297
QY 294 LFGHETTASTATSLMPLGLNTEVVOKREVEQKVEKMGMTTPKGLSMELLDQIKYTG 353
DB 294 LFGHETTASATSLITVIGLYPHVLQKVRERIKSGILCKSNQNKMDMETLEQIKYTG 357
QY 354 CVIKETELINPVPVGGFVALKTEFLNGVQIPKGNVVIYSTDTHDVADVPPNKEEFOPE 413
DB 358 CVIKETELINPVPVGGFVALKTEFLNGVQIPKGNVVIYSTDTHDVADVADIFNKEEFND 417

```

```

QY 414 REMSKGLEDSRFNVIIPRGSGRMCKYKFAKVLKIFVELTOHNMILSNPPYKMG 473
DB 418 RFIYPHPEDASRFSPFIPGGGLRSCVGEKFAKILKIFVELARHCWQJLGNPPYKMG 477
QY 474 PTIYPVDNLPTKFTSY 489
DB 478 PTIYPVDNLPTKFTSY 493

RESULT 6
US-08-622-166A-2
; Sequence 2, Application us/08622166A
; Patent No. 5952545
GENERAL INFORMATION:
APPLICANT: KONCZ, CSABA
APPLICANT: MATHUR, JAIDEEP
APPLICANT: SZEKERES, MIKLOS
APPLICANT: ALTMANN, THOMAS
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING CYTOCHROME
TITLE OF INVENTION: P450-TYPE PROTEINS INVOLVED IN THE BRASSINOSTEROID SYNTHESI
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESS: Birch, Stewart, Kolasch & Birch, LLP
STREET: P.O. Box 747
CITY: Falls Church
STATE: Virginia
COUNTRY: USA
ZIP: 22040-0747
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/622,166A
FILING DATE: 27-MAR-1996
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: SVENSSON, LEONARD R.
REGISTRATION NUMBER: 30,330
REFERENCE/DOCKET NUMBER: 0147-0153P
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 205-8000
TELEFAX: (703) 205-8050
TELEX: 248345
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 472 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-622-166A-2

Query Match
Best Local Similarity 21.6%; Score 552.5; DB 2; Length 472;
Matches 145; Conservative 98; Mismatches 205; Indels 37; Gaps 14;

QY 19 LFLIAVAKLMEMLRVDPNCRSPPLPGTGLPIGETIOL-----LQRRKFLMKRQ 73
DB 7 LLLSSIAAGFLILKRIYR--RMGLPESGLPLGTFETLIGAYKTENPPEFIDERYA 65
QY 74 KYGCIYKTHLFGNPTVRVAGADVNRQIILGEHKLVSQMPASVRTILGSDTLNVHGVQ 133
DB 66 RYGSVFMTLHFGEPITFSADPETNRFVLQNECKLPECSYPACINCLGHSILMKSGIH 125
QY 134 KKKKAKIRAFSRDL--EHTIPVIOQEVKSAIQEWLQKDSCVLVYPEKKIMFRIAMT 191
DB 126 K-RMHSILTMSFANSSIIDHMLDIDRLVRFNIDSW---SSRVLAMEBAKKTFFELTVQ 181
QY 192 ILGEP---EQIKTDEQLEAFEMIKNLFSPLIDVPFSGLYR--GLRARVFIHSKIEE 246
DB 192 LMSFPDGMSESLRKEYLIVIGF-----FSLPLPL-FSTTYKAKAIOARRKVAEALTV 233

```


QY 185 FRAMRLLGFEEBQITDEQELVEAFEE--MIKNLFLPIDVPE--SGLYRGLAR----- 237
 Db 187 YRAAFGI-----SHEDDEFVAILQERSOLFAGAFNIDFIPMLKWPQGIIVRLNKA 239
 QY 238 -----NFHSKIEENIRKKIODDNEEQKYKDALOLLIEENSRSDPEFSLQAM----- 287
 Db 240 RGALDFIDILIDHDHIOKSGKSEEVDTDVAVDLLAFYGEAKVSESDLDQNSIKILKDN 299
 QY 288 -EATTELLGHEHTTASTATSLVMTFLGNTVEVOKREVEQVEKMGMTTPGGLSMELL 346
 Db 300 IKAIMVMTGGTTVASAIEWATELKSPEDLKKVOQLAVVAGLDREVEED-----F 354
 QY 347 DQKLYGCVIKETLRINPVPGGFVRLAKTFEELNGYQIPKGMVYSICDTHVADVEPN 406
 Db 355 EKLTLYKCVLKEVYLRHLPILPLHETAEDEAVGYIIPAKSRVMIACAIGRDKNSMAD 414
 QY 407 KEFQPERENSKLED--GSRNVIYPPGGSRMCVGEKFAKVLKTLFVELTQHC--NMIL 463
 Db 415 PDRFRSRLKDGVPDFKGNNEFEIPFGSGRSCPGMQLGLVLALETTVAHL--HCFWEL 473
 QY 464 SNG 466
 Db 474 PDG 476

RESULT 9
 US-08-201-118-3
 ; Sequence 3, Application us/08201118
 ; Patent No. 5786191
 ; GENERAL INFORMATION:
 ; APPLICANT: GOLDSTEIN, Joyce A.
 ; APPLICANT: ROMKES-SPARKS, Marjorie
 ; TITLE OF INVENTION: CLONING AND EXPRESSION OF COMPLEMENTARY
 ; TITLE OF INVENTION: DNAs FOR MULTIPLE MEMBERS OF THE HUMAN CYTOCHROME P450 2C
 ; NUMBER OF SEQUENCES: 444
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Townsend and Townsend Kourlie and Crew
 ; STREET: 379 Lytton Avenue
 ; CITY: Palo Alto
 ; STATE: California
 ; COUNTRY: US
 ; ZIP: 94301
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/201,118
 ; FILING DATE: 22-FEB-1994
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/864,962
 ; FILING DATE: 09-APR-1992
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Lledeschuetz, Joe
 ; REGISTRATION NUMBER: 37,505
 ; REFERENCE/DOCKET NUMBER: 15280-192-1
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (415) 326-2400
 ; TELEFAX: (415) 326-2422
 ; INFORMATION FOR SEQ ID NO: 3:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 490 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-201-118-3

Query Match 13.2%; Score 339.5; DB 1; Length 490;
 Best Local Similarity 26.9%; Pred. No. 2,9e-25;
 Matches 139; Conservative 83; Mismatches 218; Indels 77; Gaps 22;

QY 5 TLMTFELCTIVLPVILAAVKLMEMLIRVDPNCSPPLPGPTMGIPFGITLQ--ILQ 63
 Db 3 SLVIVLYC-----LSCILLLS---LMR-----OSSGKGKLPFGFTPLPVIGNIIQIGKD 49
 QY 64 RRRFLMRKQKQYCYKTHLFGNPTVYMGADVNRQIL--LGEHKLVSVOMPASVFTILG 121
 Db 50 ISKSLTNLSKVYVPVFLYGLKPLIVLHGYEAVKEALIDLGEFSGRGIFPLAERANRG 109
 QY 122 SDTLSNVHGQVHNKKK--ALMRAFSRDLAEHYIPVIOQEVKSAIQEMLQKQDCVLYPEM 180
 Db 110 -----FGIVFSGKMKWKEIRRSMTLKNF-----GMGRSIEDRVOEARCVEELR 157
 QY 181 KLIMFRAMRLLGFEEBQI-----KTDEQ--ELVEAFEEKIK-----N 217
 Db 158 KTVASPCDDPFILIGCAPCNVICSIIIFHKRDYKQDQFLNMEKLNENIKILSSPWQICN 217
 QY 218 LBSLPIIDVPSSGLYRGIRARN--FHSKIEENIRKKIODDNEEQKYKDALOLLIEENSR 275
 Db 218 NFSPIIDY--FPGTHNKL--LKNVAFMKSYILEKVEHQESMDMNNPODFIDCFIMKKEKEK 275
 QY 276 RSD-EPEFSLOAMKEAPDELLFGGHETTASTATSLVMTFLGNTVEVOKREVEQVEKMG 334
 Db 276 HNPSEFTTESLENTAVDLFGAGCTETSTILRYALILLKHPVETAKVOEELERV----- 331
 QY 335 YYPGKGLS--MELLQIKYTGCVIKETLRINPVPGGFVRLAK--TEELNGYQIPKGMVYI 391
 Db 332 ---GNRSPCOMQDRSHMPTDAVHVEQRYLDLPLSLPHAVYCDIKFNNYILPKGTITL 388
 QY 392 YSICDT-HVADVFPKKEFQPERFMSKLEDSGRN---YIPFGGSRMCVGEKFAK 446
 Db 389 ISLTSVLHNKKE--FPPNEMFDPHHF---LDREGNKKSKRYPFSAGRICVGEALAGM 443
 QY 447 LKTLFVELTQHCNMILSNQPTMTGPTIYVNDLP 483
 Db 444 ELFLPILSTILQNFNLKSLVDPRKLDITTPVYNGFASVP 480

RESULT 10
 US-08-201-118-9
 ; Sequence 9, Application us/08201118
 ; Patent No. 5786191
 ; GENERAL INFORMATION:
 ; APPLICANT: GOLDSTEIN, Joyce A.
 ; APPLICANT: ROMKES-SPARKS, Marjorie
 ; TITLE OF INVENTION: CLONING AND EXPRESSION OF COMPLEMENTARY
 ; TITLE OF INVENTION: DNAs FOR MULTIPLE MEMBERS OF THE HUMAN CYTOCHROME P450 2C
 ; NUMBER OF SEQUENCES: 44
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Townsend and Townsend Kourlie and Crew
 ; STREET: 379 Lytton Avenue
 ; CITY: Palo Alto
 ; STATE: California
 ; COUNTRY: US
 ; ZIP: 94301
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/201,118
 ; FILING DATE: 22-FEB-1994
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/864,962
 ; FILING DATE: 09-APR-1992
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Lledeschuetz, Joe

Wed Nov 7 09:29:10 2001

us-09-668-482-2.ra1

Page 7

```

? REGISTRATION NUMBER: 37,505
? REFERENCE/DOCKET NUMBER: 15280-192-1
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (415) 326-2400
? TELEFAX: (415) 326-2422
? INFORMATION FOR SEQ ID NO: 9:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 490 amino acids
? TYPE: amino acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: protein
? OS-08-201-118-9

```

Query Match 13.2%; Score 339.5; DB 1, Length 490;
Best Local Similarity 26.9%; Pred. No. 2.9e-25;
Matches 139; Conservative 83; Mismatches 218; Indels 77; Gaps 22

[illegible]

RESULT 11
 US-08-238-821B-3
 ; Sequence 3, Application US/08238821B
 ; Patent No. 5912120
 ; GENERAL INFORMATION:
 ; APPLICANT: GOLDSTEIN, Joyce A.
 ; APPLICANT: ROMKES-SPARKS, Marjorie
 ; APPLICANT: DE MORAYS, Sonia M.F.
 ; TITLE OF INVENTION: CLONING, EXPRESSION AND DIAGNOSIS OF HUMAN
 ; TITLE OF INVENTION: Cytocrome P450 2C19: THE PRINCIPAL DETERMINANT OF S
 ; TITLE OF INVENTION: MEPHENTHOL METABOLISM
 ; NUMBER OF SEQUENCES: 61
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Townsend and Townsend and Crew LLP
 ; STREET: Two Embarcadero Center, 8th floor
 ; CITY: San Francisco
 ; STATE: California

COUNTRY: US
 ZIP: 94111
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/238,821B
 FILING DATE: 06-MAY-1994
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/201,118
 FILING DATE: 22-FEB-1994
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/864,962
 FILING DATE: 09-APR-1992
 ATTORNEY/AGENT INFORMATION:
 NAME: Liebeschuetz, Joe
 REGISTRATION NUMBER: 37,505
 REFERENCE/DOCKET NUMBER: 15280-192110US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (650) 326-2400
 TELEFAX: (650) 326-2422
 INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 490 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-238-821B-3

Query Match	13.2%;	Score 339.5;	DB 2;	Length 490;
Best Local Similarity	26.9%;	Pred. No. 2.9e-25;		
Matches 139;	Conservative 83;	Mismatches 218;	Indels 77;	Gaps 22;

[illegible]

Db 444 ELFLFLSLIOFNKLSLVDPKMLDTTPVNGFASVP 480

RESULT 12

US-08-238-821B-9
Sequence 9, Application US/0823821B
Patent No. 5912120

GENERAL INFORMATION:

APPLICANT: GOLDSTEIN, Joyce A.
APPLICANT: ROMKES-SPARKS, Marjorie
TITLE OF INVENTION: CLONING, EXPRESSION AND DIAGNOSIS OF HUMAN
TITLE OF INVENTION: CYTOCHROME P450 2C19: THE PRINCIPAL DETERMINANT OF S-
NUMBER OF SEQUENCES: 61
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, 8th floor
CITY: San Francisco
STATE: California
COUNTRY: US
ZIP: 94111

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/238,821B
FILING DATE: 06-MAY-1994
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/201,118
FILING DATE: 22-FEB-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/864,962
FILING DATE: 09-APR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Liebeschuetz, Joe
REGISTRATION NUMBER: 37,505
REFERENCE/DOCKET NUMBER: 15280-192110US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 326-2400
TELEFAX: (650) 326-2422
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 490 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-238-821B-9

Query Match 13.2%; Score 339.5; DB 2; Length 490;

Best Local Similarity 26.9%; Pred. No. 2.9e-25; Mismatches 218; Indels 77; Gaps 22;

Matches 139; Conservative 83; LMR-----QSSGKGRKPPPTPLPVIGNLTGIND 49

5 TLMWTELCITVLEFLAVKLEMLMTRVPCNSPLPPTMGLPFTGETIQL-ILQ 63

3 SLVVLVC--LSCILLLS---LMR-----QSSGKGRKPPPTPLPVIGNLTGIND 49

64 RRFPLRRKRRKRGYCYTHLFGNPTVAVMGADNVKQIL--LGHKLVSQVMPASVTIIG 121

50 ISKSLTNLSVYGVVFLYGLRPIYVLLHGYEAVKALIDLGEFSGRGIFPLAERANG 109

122 SDLSNVHGYOHKKKK-ALMRAFSRDALHETPRVIOEYKSAIDPMLQDSCLVLYPM 180

110 -----FGIVFNSGKKKKIRFRSLMTLRF-----GKKRSIEDRVQDEACLYEELR 157

181 KLMFRIAMRILLGFEPQI-----KTDEQ--ELVEAFEMIK-----N 217

Db 158 KTKASPCDPTFLIGCAPCNVICSIIFHKRFDYKDOQLNLMKLNENKILISPMIOCN 217

QY 218 LFSLPIDVPFSGLYNGLAARN--FIHKEIENIKKTIODDNEDEQYKDALOLLINSR 275

Db 218 NSSPIIDY-FPGTHNKL-LKNAVEKKSYILKRVKHOESMMNNPOPIDEFLLKMKEREK 275

QY 276 RSD-EPSFLQAKKEATBELFGHETATSTSLMFLGLTEVVOKREVEQKVEGM 334

Db 276 HNPGEFTLESLENTAVDLFGAGTETSTTLRYALLLLKRPETAKVQEEIEVI---- 331

QY 335 YTPGKGLS--WELDLQKYTCYIKETLRINPPPGGRVALK-TFELNGQYPKGNVI 391

Db 332 --GNRRSPCQDRSHMPYDAVHVEQRYIDLPTSLPHAVTCDIKFRNLTKGTTL 388

QY 392 YSICDT-HDVADVPEPKKEPEFERMSKLGDSRNF---YIPFGGSRMCYGEKFAV 446

Db 389 ISLVSLVDNKE-FENPMPDPHNF---LDGGNFKSKYEMFSGKRICVGEALAGM 443

QY 447 LKIFLVLELTHCHNWLNSGPTMKTGPTTYVDNLP 483

Db 444 ELFLFLSLIOFNKLSLVDPKMLDTTPVNGFASVP 480

RESULT 13

PCT-US95-05744-3
Sequence 3, Application PC/TUS9505744
GENERAL INFORMATION:

APPLICANT: GOLDSTEIN, Joyce A.
APPLICANT: ROMKES-SPARKS, Marjorie
TITLE OF INVENTION: CLONING, EXPRESSION AND DIAGNOSIS OF HUMAN
TITLE OF INVENTION: CYTOCHROME P450 2C19: THE PRINCIPAL DETERMINANT
NUMBER OF SEQUENCES: 61
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Khourie and Crew
STREET: 379 Lytton Avenue
CITY: Palo Alto
STATE: California
COUNTRY: US
ZIP: 94301

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/05744
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/238,821
FILING DATE: 06-MAY-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/201,118
FILING DATE: 22-FEB-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/864,962
FILING DATE: 09-APR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Dow, Karen B.
REGISTRATION NUMBER: 29,684
REFERENCE/DOCKET NUMBER: 15280-192-1-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 326-2400
TELEFAX: (415) 326-2422
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 490 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein

PCT-US95-05744-3

Query Match 13.2%; Score 339.5; DB 5; Length 490;
Best Local Similarity 26.9%; Pred. No. 2.9e-25;
Matches 139; Conservative 83; Mismatches 218; Indels 77; Gaps 22;

QY 5 TLMTVFLCTVTVLVLFLAVALKLEMLMTRVDPNCRSPPLPGTGMGLPIGTQL-ILQ 63
DB 3 SLVVLVIC---LSCILLIS---LWR-----QSSGRGLPGPPPLPVIGNILDIGIND 49
QY 64 RRKFLMKRQKQYCIKTHLFGNPTRVWAGADVROIL--LGEHKLVSQMPASVYTIIG 121
DB 50 ISKSLNLSKVGPVTVLFGKLPVVLHGVEAVKEALIDLGEESGRCGIFPLAEARANG 109
QY 122 SDPLSNVHGVOHKKKK-ALMRAFSDALEHYIPIYIOEVKSAIOEMLQKSCVLYPEM 180
DB 110 -----FGIVFNGKKWKKEIRFSLMTLRNF-----GKGKRSIEDRVQEARCLVEELR 157
QY 181 KKLMMFRIAMRILIGFEPEQI-----KTDEQ--ELVEAFEEEMIR-----N 217
DB 158 KTRASPCDPFTLIGCAPCNVICSIIFHKKRPDYKDQOFLNMEKLNENIKILSSPMIQLN 217
QY 218 LFSPLIDVPFSGLYRGRLARN--FIHSKIEENIRKKIODDNEEQKYDALQLIENSR 275
DB 218 NFSPIIDY-PPGTHNKL-LKNVAFMKSYILEKVEHQSMDMNNPDQFIDCFLMKMEKEK 275
QY 276 RSD-EPSLOAMKEAATELLFGSHETTASTATSLVMFLGINTEVOKVREEVQEKYEMGM 334
DB 276 HNPSEFTTIESLENTAVDLFGAGTETSTLRVALLLLKHPEVTAKVQEEIERVI---- 331
QY 335 YTPGKGIS--MELLQDKTYGCIKELRLNPPVPGFRAKL-TPELNGYOIPKGMNVI 391
DB 332 ---GRNSPCMDRSHMPYDAVAVHEVORYLDDLPSLPHAVTCDIKERNYILPKGTTIL 388
QY 392 YSICDT-HDVAADVPPNKEEPPERFMSKGLDEGSFRN---YIPFGGSRMCVGKEFAKY 446
DB 389 ISLSVYLDHKE-PPNPEMDPHNF-----LDEGNGFKKSKYFMPFSAGKRICVGEALAGM 443
QY 447 LKIFLVELTQHCNWLISNGPPTKGTPTIYVVDNLP 483
DB 444 ELFLFLTSILQNFMLKSLVDPKNDTTPVVGFAVSP 480

RESULT 14
PCT-US95-05744-9
; Sequence 9, Application PC/7UUS9505744
; GENERAL INFORMATION:
; APPLICANT: GOLDSTEIN, Joyce A.
; APPLICANT: ROMKES-SPARKS, Marjorie
; APPLICANT: DE MORAIS, Sonia M.F.
; TITLE OF INVENTION: CLONING, EXPRESSION AND DIAGNOSIS OF HUMAN
; TITLE OF INVENTION: CYTOCHROME P450 2C19: THE PRINCIPAL DETERMINANT
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Townsend and Townsend Kourile and Crew
; STREET: 379 Lytton Avenue
; CITY: Palo Alto
; STATE: California
; COUNTRY: US
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/05744
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/238, 821

FILING DATE: 06-MAY-1994
PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/201,118
; FILING DATE: 22-FEB-1994
PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/864,962
; FILING DATE: 09-APR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Dow, Karen B.
; REGISTRATION NUMBER: 29,684
; REFERENCE/DOCKET NUMBER: 15280-192-1-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 326-2400
; TELEFAX: (415) 326-2422
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 490 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US95-05744-9

Query Match 13.2%; Score 339.5; DB 5; Length 490;
Best Local Similarity 26.9%; Pred. No. 2.9e-25;
Matches 139; Conservative 83; Mismatches 218; Indels 77; Gaps 22;

QY 5 TLMTVFLCTVTVLVLFLAVALKLEMLMTRVDPNCRSPPLPGTGMGLPIGTQL-ILQ 63
DB 3 SLVVLVIC---LSCILLIS---LWR-----QSSGRGLPGPPPLPVIGNILDIGIND 49
QY 64 RRKFLMKRQKQYCIKTHLFGNPTRVWAGADVROIL--LGEHKLVSQMPASVYTIIG 121
DB 50 ISKSLNLSKVGPVTVLFGKLPVVLHGVEAVKEALIDLGEESGRCGIFPLAEARANG 109
QY 122 SDPLSNVHGVOHKKKK-ALMRAFSDALEHYIPIYIOEVKSAIOEMLQKSCVLYPEM 180
DB 110 -----FGIVFNGKKWKKEIRFSLMTLRNF-----GKGKRSIEDRVQEARCLVEELR 157
QY 181 KKLMMFRIAMRILIGFEPEQI-----KTDEQ--ELVEAFEEEMIR-----N 217
DB 158 KTRASPCDPFTLIGCAPCNVICSIIFHKKRPDYKDQOFLNMEKLNENIKILSSPMIQLN 217
QY 218 LFSPLIDVPFSGLYRGRLARN--FIHSKIEENIRKKIODDNEEQKYDALQLIENSR 275
DB 218 NFSPIIDY-PPGTHNKL-LKNVAFMKSYILEKVEHQSMDMNNPDQFIDCFLMKMEKEK 275
QY 276 RSD-EPSLOAMKEAATELLFGSHETTASTATSLVMFLGINTEVOKVREEVQEKYEMGM 334
DB 276 HNPSEFTTIESLENTAVDLFGAGTETSTLRVALLLLKHPEVTAKVQEEIERVI---- 331
QY 335 YTPGKGIS--MELLQDKTYGCIKELRLNPPVPGFRAKL-TPELNGYOIPKGMNVI 391
DB 332 ---GRNSPCMDRSHMPYDAVAVHEVORYLDDLPSLPHAVTCDIKERNYILPKGTTIL 388
QY 392 YSICDT-HDVAADVPPNKEEPPERFMSKGLDEGSFRN---YIPFGGSRMCVGKEFAKY 446
DB 389 ISLSVYLDHKE-PPNPEMDPHNF-----LDEGNGFKKSKYFMPFSAGKRICVGEALAGM 443
QY 447 LKIFLVELTQHCNWLISNGPPTKGTPTIYVVDNLP 483
DB 444 ELFLFLTSILQNFMLKSLVDPKNDTTPVVGFAVSP 480

RESULT 15
US-09-091-432-2
; Sequence 2, Application US/09091432
; Patent No. 5981837
; GENERAL INFORMATION:
; APPLICANT: Chapple, Clint
; TITLE OF INVENTION: A Method for Regulation of Plant Lignin Composition
; FILE REFERENCE: 7024-325

```
; CURRENT APPLICATION NUMBER: US/09/091,432
; EARLIER FILING DATE: 1998-06-18
; EARLIER APPLICATION NUMBER: PCT/US96/20094
; EARLIER FILING DATE: 1996-12-19
; EARLIER APPLICATION NUMBER: US 60/009,119
; EARLIER FILING DATE: 1995-12-22
; EARLIER APPLICATION NUMBER: US 60/013,388
; EARLIER FILING DATE: 1996-03-14
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: Microsoft Word 2.0C
; SEQ ID NO 2
; LENGTH: 520
; TYPE: PRF
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: Sequence is deduced from DNA sequence of SEQ ID NO:1
; Patent No. 5981837
US-09-091-432-2
```

```
Query Match      12.9%; Score 330.5; DB 2; Length 520;
Best local Similarity 24.9%; Pred. No. 2,5e-24;
Matches 125; Conservative 89; Mismatches 207; Indels 81; Gaps 20;

QY 13 TIVLPVLLFLAAYKLMEMIRRYDPCNSPLPPGTMLPFISGTTQLLIQRRKFLMKR 72
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 18 SLIVVSLF-----IFISFTRRRRPYPGPGWPIIGMLMDQITHRGLANLA 68
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 73 OKYK--CIYK---THFGNPTVRVMGADNVRQIILGEHKLVSVOVPASVR---TILGSD 123
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 69 KTYGGLCHLMGFLHMA-----VSSPEVARQYLVQYDVSFSNR-PATIAISLYITDRAD 122
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 124 TLSNVHG-VQHKKKKAIMAFSRLALEHTIPYIQDEVKSAIQEWLOKSC-----VLVY 177
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 123 MAFAHYGPFWRMKVCVMVFSKRRAESW----ASVRDEVDMKVRVSCNVGKPLNVG 177
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 178 PEMKKLMEFLAMRILILGFEPEQITDEQLVEAFEEMLK--NLFSLPIDVPFSGLY---- 231
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 178 EQIFALTRNTTYRAAFSACEK---GGDEFIRILQESKILFGAFNVADFIPIFGWIDPDG 234
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 232 ---RGLAARN---FTHSKTEENIRKK-----IODDNEENQKTKDALQLLIENSRSDE 279
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 235 INKRLVAKARDLDGFIIDIDEHMKKENONAVDDGDVDTDMVDLLAFYSEAKLYSE 294
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 280 PFSIQ-----AMKPEATELLFGHETASTATSLVMFLGINTEVVOKVREVEQKVE 331
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 295 TADIQNSIKTLRQIKAIIMDVMMGTETVYASALEMALTLELRSPEDLKRVOQLAEVY- 353
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 332 MGMTTPGKGLSMEL---LDQLKTYGCVIKETLRINPPVPGGFRVALKTFELNGYQIPKG 387
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 354 -----GLDRRVEESDIDKLTLYLCKTLKELTMRHPPIPLLHETAEPTSIDGFFIPKK 405
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 388 MNVYISICDTHDVAVDPNKEEPOPERFMSKGLSD--GSRFNITPFGGSRMCVGKEFAK 445
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 406 SRVIMNFAIGRDPSTWTDPTFRPSRFLBPGVDFDEKGSNFEFLIPGSGRRSCPGMQLGL 465
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 446 VLKRIPLVELQHC-NWILSNG 466
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 466 YALDLAVAHIL-HCFTWKLPDGL 486
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

Search completed: November 6, 2001, 13:27:04
Job time: 9917 sec


```

QY 343 MELLDOQKTKGVIKETLLKINPNYPVGGFVAKLTPELNGYQIQPKGMNVIYSICDTHVDAD 402
Db 295 AETLKMKPILDDYLEDEVLILIPVGGGFRRELIDQCFQGFHPKGMIVSYQISQTHADDP 354
QY 403 VEPNKEEFPQPEFFMSKGLDGSR-----FNVIPEFGGSGMCGKEFAKVLITFVELVQ 457
Db 355 LVPDEKFPDEPF-----TFDGSATNHPRAHYPPFGGGLRECLGKEFAKLEKFAETRLIQ 410
QY 458 HCNWILSNGP-TMKTGPTIYVDNLPTKETS 488
Db 411 QPDMWLLPQNLLELVTPSPRPKDNILRVKLHS 442

RESULT 2
cytochrome P450 CYP90 - Arabidopsis thaliana
N:Contains: oxidoreductase (EC 1.-.-.-)
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C:Accession: S55379
R:Sezkeres, M.; Nemeth, K.; Koncz, Z.; Nagy, F.; Koncz, C.
submitted to the EMBL Data Library, May 1995
A:Reference number: S55379
A:Accession: S55379
A:Molecule type: mRNA
A:Residues: 1-472 <SZEP>
A:Cross-references: EMBL:X87367; NID:g853718; PIDN:CAA60793.1; PID:g853719
C:Genetics:
A:Gene: CYP90
C:Superfamily: Synechocystis cytochrome P450 slr0574; cytochrome P450 homology
C:Keywords: chromoprotein; electron transfer; home; iron; metalloprotein; monooxygenase.
F:275-440/bomain: cytochrome P450 homology <F45>
F:418/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 21.6%; Score 552.5; DB 1; Length 472;
Best Local Similarity 29.9%; Pred. No. 2e-29;
Matches 145; Conservative 98; Mismatches 205; Indels 37; Gaps 14;

QY 19 LLFLAAVLMEMLMIRVDPNCRSPDLPGTGMGLPEIGTOLI-----LQRRKFLMRQ 73
Db 7 LLLLSIAGFLLLRTRRFR-RMGLPPSGSLGPLIGETPOLIGAYKTENPEPFIDERYA 65
QY 74 KYGGIYKHLNGNPVRMGADNRQQLLGEHKLVSYQWPAVSRIILGSDPLSNVHCYQH 133
Db 66 KYGVFMFLHGEPTIISADPEINFEVLQNGEKLFECSYPASICNLGKSHLLMKGSLH 125
QY 134 KKKKAIWRAFSPDAL--EHYTPYIQOEKSAIOEWLQKDCSVLYVPMKILMIRARI 191
Db 126 K-RMHSILTMSFANSIITKDHMLDIRVRENLDW---SSRVLLMEAKKITTELYKQ 181
QY 192 LLGEP-----EQITDEQLEVAEEEMIKNLFSLPIDVPSGLYR-GIRANPIHSKITEE 246
Db 182 LMSDPPGWSLSLKEVLLYIEGF-----FSLPLP-LFSTYRKAIQARAKVAEALTV 233
QY 247 NIKRIIDDDNENOKKDALQLLIENSRSRDEPFSIQAMEATTELLFGHETASTAT 306
Db 234 VYMKR-KEEEEGEAKRKMIALLL-----AADDGFSDEIYDELVALLVAGYETTSITMT 288
QY 307 SLVNFLLGNTENVOKVREVEQEKVEMGYTPGKLSMELLDQLKYTGCVIKETLRINPV 366
Db 289 LAVFLETETPLATLQLEE-HEKIR-AMKSDSYSELEMSDYKSMPTQCVVNETLRVANI 346
QY 367 PGGFRVALKTFELNGYQIQPKGMNVIYSICDTHVDADVPNKEEQPERFMSKGLDGSRF 426
Db 347 GGVRFRMTDVEIKGYIPIKGMKVFSSFRAYHLDPNHFKARIPNPRWQSNSTYTPSPN 406
QY 427 NYIPFGGSGMCGKEFAKVLITFVELTQHCNWIILSNGPMTKKTGPTIYVDNLPTKF 486
Db 407 VEPNKEEFPQPEFFMSKGLDGSR-----FNVIPEFGGSGMCGKEFAKVLITFVELVQ 457
QY 487 TSIVR 491

```

```

Db          462  PIRVK  466
:|:
RESULT      3
A84859
probable cytochrome P450 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 23-Mar-2001
C:Accession: A84859
R:Lin. X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Beaulieu, M.I.; Town, C.D.; Fujii,
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tal-
leus, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; V
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana
A:Reference number: A84420; MUID:20083487
A:Accession: A84859
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-485 <STO>
A:Cross-References: GB:AE002093; NID:g4512670; PIDN:AAD21724.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g42850
A:Map position: 2
C:Superfamily: Synchocystis cytochrome P450 slr0574; cytochrome P450 homolog
C:Keywords: heme; iron; metalloprotein
F:432/Binding site: heme iron (Cys) (axial ligand) #status predicted

```

Query Match	20.8%	Score 532.5	DB 2:	Length 485;
Best Local Similarity	28.2%	Pred. No. 4.5e-28;		
Matches 143;	Conservative 101;	Mismatches 214;	Indels 49;	Gaps 15;
QY 4 YLWMTFLC--TIVLPVLEFLAIVKEMLMARRVDPCNSPLPQGTMLPFGTGLTI 61				
DB 7 FVLMSVPLCIATISTSTLEFFR-KKHNFILKKIÖK-KKLLDGEKMLPIGEMMDY 62				
QY 62 LÖRK-----FLNRKÖKYGCITYTHLFGNPTVRYMGADNVÖLLGSHKLVSVOMPAS 115				
DB 63 KÄÖKSNRVFEEDVNDRIIKHGNIETKRMGSPTIYVNGEANRLLILNSFSLVSVSPSS 122				
QY 116 VTIILGSDTLNVHGVÖHKKKALIMRAFSPDALIEHIVPDIÖEYKSAIÖ-EMLÖKDSCV 174				
DB 123 SVÖLGMGNMCIMAKÖGEKRIYLRGIYANSLSYIGLESILPKLÖDYKFNHETMRÖKE-I 181				
QY 175 LVYPRMKKLMRIAMRIILGEPEDÖIKTDEÖELVAFEMIKNLSPLIDVPSGLYGL 234				
DB 182 SLYSRAKVLFTFYVEECYLG-----IKY-ETGMLEVFERYELGVALPIEFPCSFARAK 235				
QY 235 RARNIHSKIEENIRKKTÖDDNENEÖKYKDAL-ÖLLIENSRRSDPFSIÖAMKKAATP- 292				
DB 236 KARLEIEFFIVGVAKVERKREMEKBEAKERNPTLFSRLYE-----ELIKVITIEE 284				
QY 293 -----LLEGGHTTSTATSLVMEIÖGLNTEYVÖK-VREÖVÖKVENKMTPEKGLISM 343				
DB 285 EVDNMVLLVPAAHDTTSTAYAMSMTFKMLAÖHPTCRDLOJÖEHAÖTKANKG--EÖEYLLTV 341				
QY 344 ELLDÖLKTGVCVIKETTILNPVPÖGFAVAKTEFLNÖYÖIPRGKNVYVSIÖDTHDADV 403				
DB 342 EDVAKMKMSQVÖVRETRMTLSPRIÖGSPFKAAVIDYÖGTYTIPRKGMKILMTYIGTHYNEI 401				
QY 404 FPNKEEÖPÖEPFMKSGLEDGSRFNTIIPRGSGSRMCVGEKPAKVLKTLVLETHÖCNMIL 463				
DB 402 FÖDPASÖPPTPF-KPÖI--AYTYLPEGGÖPRICAGHÖLAKISILVYMHVYVÖGFWDSL 457				
QY 464 SNGPPTMTKGTPTIYVDNLPTKFTSYV 490				
DB 458 VYPEÖTISMDPLPEPSSGMPTIKISPKV 484				


```

Query Match      17.8%; Score 455; DB 2; Length 457;
Best Local Similarity 28.6%; Pred. No. 6.1e-23;
Matches 130; Conservative 90; Mismatches 185; Indels 50; Gaps 16;

OY 44 LPRGMGLPFIETGLIQLL-----QRRKFLMKRQKQCYIKTHLFGNPYRVAGDANVR 98
   : : : : : | | | | | : : : : : | | | | | : : : : : | | | | |
Db 2 IPNGLGMPVIGETLNFACGYSSRPYEMDKRSLGKYEKVTNIGFTIISTDAE-VN 60

OY 99 OLLGEHKLIVAV-QMPASVRTILGSDTLNNGVQHKNNKKAIRAFSRDALHYIPIYIQ 157
   : : : : : | | | | | : : : : : | | | | | : : : : : | | | | |
Db 61 KVLNLONHNTVPAPAKRTITELGNSLISLNG-PRKRLKTLTGAFIURS--PHLKORIT 117

OY 158 QEVKSAI-----QEWLQKDCVLYVPEMKKLMRIAMRIILGFEPEQIKTDEQLEAFEE 213
   : : : : : | | | | | : : : : : | | | | | : : : : : | | | | |
Db 118 RDIKESAVVLTIASMAQL-PLVHVODEIKKMTFELIVKLMSTSGE---DNKILKEPEE 173

OY 214 MKNLFSLPIDVPFSGVGLRGRLARNFHISKIENIRKKIODDD--NENEQYKDALOLLI 271
   : : : : : | | | | | : : : : : | | | | | : : : : : | | | | |
Db 174 FIKGLICIPIKFPGTRILKSLKAK-----ERLIKVKYKVEERQVAMTTSTANDVIVDL 229

OY 272 E---NSRRSDPEPSLOAMKEATELLFGGHETTASTATSLVMTGLNTEVYQKVEEVQE 328
   : : : : : | | | | | : : : : : | | | | | : : : : : | | | | |
Db 230 RCGGSEKQSQSPDFVSK--IVEMPIGEEPTAMTLAVKFLSDNVALAKL---VEE 284

OY 329 KYEMGMTPPGKGLSMELLDOIKYTC-----VIKETLRINPVPVGGFVALKTFELN 380
   : : : : : | | | | | : : : : : | | | | | : : : : : | | | | |
Db 285 NMEE-----KRRKLELGEYKMTDYMSLSTFQNYINETLRANITNGWRKALKDVEIK 338

OY 381 GVOIRKGMNVIYSICDTHVDVAFPNKKEFOPERPMSKGLSDGSRFNYIPGGGSRMCVG 440
   : : : : : | | | | | : : : : : | | | | | : : : : : | | | | |
Db 339 GLIIRGMCVLSFISVHMDEDIYINPYQFDPWRMDRINGSANSSICTPFGGGRILCPG 398

OY 441 KEPAVLKIFLVELTQHCNWLISNGP-----PTMK 471
   : : : : : | | | | | : : : : : | | | | | : : : : : | | | | |
Db 399 LELSKLEISIFLHNLVTRYSMTAEDEIYSPYTK 433

RESULT 9
H68185
hypothetical protein [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: H68185
R:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Hultzar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719
A:Accession: H68185
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-490 <STO>
A:Cross-references: GB:AE005172; NID:g2388581; PIDN:AA871462.1; GSPDB:GN00141
A:Map position: 1
C:Superfamily: Synchocystis cytochrome P450 slr0574; cytochrome P450 homology

Query Match      17.6%; Score 450; DB 2; Length 490;
Best Local Similarity 27.6%; Pred. No. 1.4e-22;
Matches 125; Conservative 84; Mismatches 220; Indels 24; Gaps 10;

OY 41 RSPLPPTGMLPFIETGLIQLR-----KFLMRKQKYG--CIYKTHLFGNPYRVWG 93
   : : : : : | | | | | : : : : : | | | | | : : : : : | | | | |
Db 42 RHYLPDGLDMPFIETGLNMLSPFLRAFKTSDPSFTRLIKRGPKGIYKAHFNPSITVTT 101

```

```

OY 94 ADVNRQTLGHEHKLIVSQMPASVRTILGSDTLNNGVQHKNNKKAIRAFS-RDALEHY 152
   : : : : : | | | | | : : : : : | | | | | : : : : : | | | | |
Db 102 SDICRRVLTDDAF-KRQWPTSMELIGKRSFYGFEEKRLRLTAAPVNGHEALSTY 160

OY 153 IPVIOQEVKSAIQEWLQKDCVLYVPEMKKLMRIAMRIILGFEPEQIKTDEQLEAFEE 212
   : : : : : | | | | | : : : : : | | | | | : : : : : | | | | |
Db 161 IPIYIEENVITVLDKWTIKMGFEFL-THLRKLTFRITIMYIFLSESENV---DALREY 216

OY 213 EMKINFSLPIDVPFSGVGLRGRLARNFHISKIENIRKKIODDNEQKY-----KDALQ 268
   : : : : : | | | | | : : : : : | | | | | : : : : : | | | | |
Db 217 ALNNGVAMAVNIPGFAYHRLAKRKTVAQFQSYTER---RNRKQITLSNKKDDMD 272

OY 269 LITNSRRSDPEPSLOAMKEATELLFGGHETTASTATSLVMTGLNTEVYQKVEEVQE 328
   : : : : : | | | | | : : : : : | | | | | : : : : : | | | | |
Db 273 NLNWKEDCGKTLDDDEITIDVLLMYLNAGHESGHTIMATVYLOEHPEVLOKAKAF-QE 331

OY 329 KVENGMTPPGKGLSMELLDOIKYTCIKETLRINPVPVGGFVALKTFELNGYQIPKGM 388
   : : : : : | | | | | : : : : : | | | | | : : : : : | | | | |
Db 332 MLKSRPEGKGLSLKETRKMFEFLSQVDETLRVITPSLTAFAEAKTDVEMNGYLLPIKGM 391

OY 389 NVYISICDTHVDVAFPNKKEFOPERPMSKGLSDGSRFNYIPGGGSRMCVGKFAVLL 448
   : : : : : | | | | | : : : : : | | | | | : : : : : | | | | |
Db 392 KVLTFMRDVIHDPEVFPDPKFPDARMDCGVPKAGAF--LPFGAGSHLCPGMDLAKET 449

OY 449 KIFLVELTQHCNWLISNGPPTMKGTPTYPVDN 481
   : : : : : | | | | | : : : : : | | | | | : : : : : | | | | |
Db 450 SIFLHHLKTKYQKRSNPECVMLPHTTRPTDN 482

RESULT 10
B84733
probable cytochrome P450 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
C:Accession: B84733
R:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487
A:Accession: B84733
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-489 <STO>
A:Cross-references: GB:AE002093; NID:g3831452; PIDN:AAC69934.1; GSPDB:GN00139
A:Gene: At2g32440
A:Map position: 2
C:Superfamily: Synchocystis cytochrome P450 slr0574; cytochrome P450 homology

Query Match      15.4%; Score 394.5; DB 2; Length 489;
Best Local Similarity 25.6%; Pred. No. 7.3e-19;
Matches 130; Conservative 96; Mismatches 227; Indels 55; Gaps 17;

OY 9 TFLCTIVLPVLLFLAAYKLMEMLRVD-----PNCSSPLPPTGMLPFIETGLQ 60
   : : : : : | | | | | : : : : : | | | | | : : : : : | | | | |
Db 4 TGLILMFPRLILGLFLKW--VLKRVNVIYYSKLGKKHNYLPDGLGMPVIGNMMSF 60

OY 61 I-----LQRRKFLMRKQKYG--CIYKTHLFGNPYRVAGDANVRQTLGHEHKLIVSQMP 113
   : : : : : | | | | | : : : : : | | | | | : : : : : | | | | |
Db 61 LRAFKTSDPESTIQSYITRIRYKGTGIYKKNMGYPCVLVTTPTETCRVLTDDAF-HIGMP 119

OY 114 ASVRTILGSDTLNNGVQHKNNKKAIRAFS-RDALEHYIPVIOQEVKSAIQEWLQKDS 172
   : : : : : | | | | | : : : : : | | | | | : : : : : | | | | |
Db 120 KSTMKLIGKRSFYGFEEKRLRLTAPVNGPEALSVYIOFTEETVNTDLEKW-SKMG 178

OY 173 CYLVVPEMKKLMRIAMRIILGFEPEQIKTDEQLEAFEEIKNL-----FSLPIDVPFS 228
   : : : : : | | | | | : : : : : | | | | | : : : : : | | | | |
Db 179 ETEFLSHLKLTFKVIYMTIFLSESEHY-----MDLEERYTNLNGVAMGNTLPGF 231

OY 229 GLYRGRLARNFHISKIENIRKKIODDNEQKYKDALOLLI-----ENSRSDPEPSLO 284

```

[illegible]

RESULT 11
H96759
Probable steroid 22-alpha-hydroxylase T9L24.44 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
A:Accession: H96759
R/Theologias: A.; Ecker, J.R.; Palm, C.J.; Federpiel, N.A.; Kaul, S.; White, O.; Alonso,
Chiu, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Hultzer, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maitl, R.; Marzilli,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719
A:Accession: H96759
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-512 <STO>
A:Cross-references: GB:AE005173; NID:g11120803; PIDN:AAG30983.1; GSPDB:GN00141
C:Genetics:
A:Gene: T9L24.44
A:Map position: 1

Query Match	14.9%	Score 382.5;	DB 2;	Length 512;
Best Local Similarity	26.6%;	Pred. No. 4.9e-18;		
Matches 130;	Conservative 76;	Mismatches 20;	Indels 83;	Gaps 14;

QY	9	TFLCTIVLPVLLFLTAAYKLMEMLMIRVDPMCRSP--LPFGMLPFIGET-----	57
		: : : :	
		: : : :	
Db	19	TEFAFIIFILLGIARRK-----RAHNR.LPPGSRGMPLIGDTPAMLNVAAG	65
		: : : :	
		: : : :	
QY	58	-----LQILQRRKRLMKRQKGYGCTYKTHLFGNPLVRYMGADN	96
		: : : :	
		: : : :	
Db	66	SHPSSEYVEKQIKKTVSLCSVLLILKRPDNGSGENEIYRGIFSCSLGKMAVVSADPDF	125
		: : : :	
		: : : :	
QY	97	VROILLGEHKLVSYOMPASVYRTILGSDPLSVNHGVQHKKKAIMRAFSPDALE-HY---	152
		: : : :	
		: : : :	
Db	126	NRFITMONEGKLFQSSYKPSFREDYVGKDGIVITHGQQRRLRSIASSMNRHQKTHFLEV	185
		: : : :	
		: : : :	
QY	153	IPVLIQGEVKSALQIOWLOKDCSVLYPEPKKIMFRIAMFILLGFEDEQIKTDEQELVEAFE	212
		: : : :	
		: : : :	
Db	186	IPVVMQLTSLNFKK-----GEVVLILQDICIORKVAIHLVMQLGVGSSSEVDENSOFLS	238
		: : : :	
		: : : :	
QY	213	EMIKNLPSLPIDVPGSGYGRGLRARNFHSKIEENIRKKIIDDODNENQKKDAQLLIE	272
		: : : :	
		: : : :	
Db	239	DFVDCGCLSPVPIDLPGFLYNNKMKARKELIKINKTIERKLO-NKAASDTAGNGVIGRLIE	297
		: : : :	
		: : : :	
QY	273	NSRRSDPFLQAMKEATELLFGGHETTASTATSLVETLGNTEVYQKREVEQYKEM	332
		: : : :	
		: : : :	

```

Db 298 -----EBSLPNBSMADETIINLLFAGNETTSKTMFLFAYFLTHCPKANTOLLEE-HDRLAG 351
QY 333 GMYTEPGKLSMELLDOLKTYGTCVAKETLRINPVPVG-----FRVALKTFELNGYOIPIKG 387
Db 352 GN-----LTMODKYTMFLTCVVIDETLR-----GGAIATMLMREAEEDYSYDYIYIPKG 400
QY 388 MNVIYSICDTHDVADEVPEENKEEFQPERFMSKGLDEGSRFN---YIPFGSGRMVCYKKE 443
Db 401 CFVVPFLSAVHLDESYSYKESLSFPMRWMLDETQOKRNMTSPFYCPFGGGRTRFCGAEI 460
QY 444 AKVLKIKFL 452
Db 461 AKQLATLFL 469

```

RESULT 12
T02263
cytochrome P450 DWARF3 - maize
N:Contains: oxidoreductase (EC 1.-.-.-)
C:Species: Zea mays (maize)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C:Accession: T02263
R:Winkler, R.G.: Helentjaris, T.
Plant Cell 7, 1307-1317, 1995
A:Title: The maize dwarf3 gene encodes a cytochrome P450-mediated early step in gibberellin biosynthesis
A:Reference number: Z14648; MUID:96004534
A:Accession: T02263
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-519 <WIN>
A:Cross-references: EMBL:U32579; NID:g987266; PIDN:AAC49067.1; PID:g987267
A:Experimental source: strain B73
C:Genetics:
A:Gene: dwarf3
C:Function:
A:Description: involved in an early step in gibberellin biosynthesis
A:Pathway: gibberellin biosynthesis
C:Superfamily: Synchocystis cytochrome P450 slr0574; cytochrome P450 homology
C:Keywords: oxidoreductase
;325-488/Domain: cytochrome P450 homology <P45>

	Query Match	14.68;	Score 375;	DB 1;	Length 519;	
	Best Local Similarity	25.68;	Pred. No. 1.6e-17;			
	Matches 119;	Conservative 94;	Mismatches 216;	Indels 36;	Gaps 13;	
QY	41 RSPLEPGMTGPIETGETLQTLILORRK-----FLRKRQKRYG--CIYKTHLGPNPVRWVG	93				
Dd	69 RARLPGEIMGPIVGGMAAFRLAFRSGCKPDFAFASFFRGTRFGTGYRRSFMSPTVLVTT	128				
QY	94 ADNRQQLLGSHKLVSVOVPASVFTILGSDFLSNVHGVQHKKKK--AIAPAESDALE	150				
Dd	129 AEGCKOYLMDDDAVYT-GMPKATVALVGRSEFVAMPDYDEHHRIKRTLAAPINGF--DALT	185				
QY	151 HYIPVIOOEYVKSALOEMLQKDSCLVUYEEMKKLMFERIMRITLLGEPPOKITDDELVEA	210				
Dd	186 GYLPEFDRTVTSSLRAMADHGGSVEFTELKRMFKIIVQLFPLG-GADQANT--RALERS	242				
QY	211 FEEMIKNLFSLPIDVPPSGLYRG-LRAENFIHSKIIEENIRKKIODDENENOKYK-----	264				
Dd	243 YTELTNYGMRAAMAILNP-GFAYRGALRRARRRLVAIVLOGLV-----DERAARAAGVSGG	294				
QY	265 --DALTQLLIENSRSRDEFSIQAKMKEATELLPBGHETTASTAISLVWFGLNLNEVOKV	322				
Dd	295 GVDMMDRILLIEQDERGRHLHDDDEITIDVLVMTLNGHESHGIITMAYTFILOENDMFARA	354				
QY	323 REEEVQEVEMGMTPYPKGLSMELLDOUKTYTCVIKETLRIMPYPVGGCFVALTKFEELNG	382				
Dd	355 KAE-QEAIMSIPSISGQGLTLDRDKMEYLSQVYIDEITRLVNISFVSRSQAIRTDROVPANGY	413				
QY	383 QIRKGNAVITYICTDHVADVFPNKKEPQEBREMSKGLIEDSGRFNYPFPGGSRMCVYKE	442				
Dd	414 LIPKGMVOLMYRSVNDHPQVYDPRTKFDSRWGCHSPRAGT---FLAFIGGARIGKND	470				

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2001, 13:34:12 ; Search time 115.92 Seconds
(without alignments)
561.543 Million cell updates/sec

Title: US-09-668-482-2

Sequence: 1 MGITYLMTVFLCTIVLPVLL.....GPTIYPVDNLPTKTSYVKN 492

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 425026 segs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
Maximum Match 1008

Listing first 45 summaries

Database :

SPTREMBL_16:*
1: sp.archaea:*
2: sp.bacteria:*
3: sp.fungi:*
4: sp.human:*
5: sp.invertebrate:*
6: sp.mammal:*
7: sp.mhc:*
8: sp.organelle:*
9: sp.phage:*
10: sp.plant:*
11: sp.rodent:*
12: sp.unclassified:*
13: sp.vertebrate:*
14: sp.virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query Length	DB ID	Description
1	1816.5	70.9	492	13	09PUB4
2	1738	67.8	497	11	09RIE4
3	1646.5	64.2	492	13	093323
4	1034.5	40.4	512	4	09NR63
5	1022	39.9	525	4	09NP41
6	689.5	26.9	444	2	059990
7	598.5	23.4	477	10	09LVY7
8	532.5	20.8	485	10	09SUH2
9	523	20.4	463	10	09FH76
10	515.5	20.1	474	10	09LKH7
11	503	19.6	482	10	081077
12	503	19.6	513	10	064989
13	503	19.6	513	10	09SCQ9
14	489.5	19.1	457	10	065624
15	481.5	18.8	465	10	09FMA5
16	474.5	18.5	465	10	09LH81
17	460.5	18.0	478	10	09LH73
18	460	17.9	107	13	09PUC2
19	455	17.8	457	10	09M066

20	455	17.8	524	10	023242	023242 arabidopsis
21	441	17.2	443	10	09LJK2	09LJK2 arabidopsis
22	435.5	17.0	466	10	09FOX4	09FOX4 cucurbita m
23	408.5	15.9	464	10	09LIC5	09LIC5 arabidopsis
24	403.5	15.7	518	10	09F138	09F138 arabidopsis
25	397.5	15.5	735	10	09LGI7	09LGI7 arabidopsis
26	394.5	15.4	489	10	09ZV72	09ZV72 arabidopsis
27	382.5	14.9	512	10	09FX29	09FX29 arabidopsis
28	370.5	14.5	504	10	09SNG3	09SNG3 oryza sativ
29	366	14.3	475	2	09X7G9	09X7G9 myxococcus
30	356.5	13.9	486	10	09LIE9	09LIE9 arabidopsis
31	356.5	13.9	495	10	09TOK2	09TOK2 arabidopsis
32	353.5	13.8	474	10	09SH7	09SH7 arabidopsis
33	350.5	13.7	491	5	09NGK3	09NGK3 tribolium c
34	349	13.6	418	10	09LVY3	09LVY3 arabidopsis
35	349	13.6	490	10	09TOK0	09TOK0 arabidopsis
36	347.5	13.6	516	10	09STI1	09STI1 arabidopsis
37	347	13.5	503	4	09HB55	09HB55 homo sapien
38	346	13.5	508	13	09PUB4	09PUB4 gallus gall
39	345.5	13.5	464	10	004949	004949 arabidopsis
40	344.5	13.4	461	2	09K498	09K498 streptomyce
41	343	13.4	486	11	09J1Y3	09J1Y3 mus musculu
42	343	13.4	503	11	09J1P8	09J1P8 mus musculu
43	343	13.4	511	10	09SWR1	09SWR1 liquidambar
44	338.5	13.2	487	4	016756	016756 homo sapien
45	336.5	13.1	504	4	09HB54	09HB54 homo sapien

ALIGNMENTS

RESULT 1
ID 09PUB4 PRELIMINARY: PRT: 492 AA.
AC 09PUB4;
DT 01-MAY-2000 (TREMBLrel, 13, Created)
DT 01-MAY-2000 (TREMBLrel, 13, Last sequence update)
DE 01-MAR-2001 (TREMBLrel, 16, Last annotation update)
DE CYTOCHROME P450 26 (BC 1.14, "-") (RETINOIC ACID-METABOLIZING
DE CYTOCHROME) (P450RA1) (RETINOIC ACID 4-HYDROXYLASE).
GN CYP26A1.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RA Swindell E.C., Thaller C., Sockanathan S., Petkovich M., Jessell T.M.,
RA Eichele G.;
RT "Complementary Domains of Retinoic Acid Production and Degradation in
RT the Early Chick Embryo."
RL Dev. Biol. 0:0-0(2000).
CC -1- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) = ROH +
CC OXIDIZED FLAVOPROTEIN + H(2)O (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC EMBL: AF199462; AAF09250.1; -.
DR InterPro: IPR001128; -.
DR Pfam: PF00067; P450; 1.
DR PRINTS: PR00385; P450.
DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
KW Electron transport; Endoplasmic reticulum; Heme; Membrane; Microsome;
KW Monooxygenase; Oxidoreductase.
SQ SEQUENCE 492 AA: 55264 MW: 7F28B72E75C322FB CRC64;

Query Match 70.9%; Score 1816.5; DB 13; Length 492;
Best Local Similarity 70.4%; Pred. No. 1.5e-112;
Matches 345; Conservative 63; Mismatches 79; Indels 3; Gaps 2;
QY 1 MGITYLMTVFLCTIVLPVLLFLAAVRLWEMLMIRVDPNCRSLPPGTGFLFGETIQL 60

```

Db 1 MGFSALVASACTFLDLLFLAAYRLMDLVCASGRDPSOCLPFGTMOGLPFGGELIOM 60
QY 61 ILORRKLPMKROKGYCYTHLFGNPTVVMGADNVROILLGHEKLVSVOMPASVETIL 120
Db 61 VLORRKLPMKRRKGYCYTHLFGNPTVVMGADNVROILLGHEKLVSVOMPASVETIL 120
QY 121 GSDPLSNVHGVQHNKKKAIKIMAFSRDALHPIVIOEVSALIOEMLOKDCVLYPEM 180
Db 121 GSGCLSNLHNGOHRKKKIVIMQAFSRDALHPIVIOEVSALIOEMLOKDCVLYPEM 180
QY 181 KLMFRLAMRILGEPBOQITD-EQELVEAFEMIKNLSLPIDVPFSGLYRGLRANF 239
Db 181 KLMFRLAMRILGEPBOQITD-EQELVEAFEMIKNLSLPIDVPFSGLYRGLRANF 239
QY 240 IHSKIEENIRKRIODDNEENOKYKDALQILLIENSRRSDPEFSLOAKKEATLLEFGHE 299
Db 240 IHSKIEENIRKRIODDNEENOKYKDALQILLIENSRRSDPEFSLOAKKEATLLEFGHE 299
QY 241 IHAKEIENIRKRIODDNEENOKYKDALQILLIENSRRSDPEFSLOAKKEATLLEFGHE 298
Db 241 IHAKEIENIRKRIODDNEENOKYKDALQILLIENSRRSDPEFSLOAKKEATLLEFGHE 298
QY 300 TTASTATSLVWFLDNLNTEVVOKREVEQVEKMGMTTPEGKLSMELLDOILKYTCVET 359
Db 300 TTASTATSLVWFLDNLNTEVVOKREVEQVEKMGMTTPEGKLSMELLDOILKYTCVET 359
QY 359 LRLSPVVGGRFALKTLELNGYQIPKGMNVTISICDTHVADVPKKEEOPREFMSKG 418
Db 359 LRLSPVVGGRFALKTLELNGYQIPKGMNVTISICDTHVADVPKKEEOPREFMSKG 418
QY 420 LEDSRNRYTFPGGSRMCVGEKFAKYLKFLVELTOHCNMLILSNPPTKGTPTYPV 479
Db 420 LEDSRNRYTFPGGSRMCVGEKFAKYLKFLVELTOHCNMLILSNPPTKGTPTYPV 479
QY 419 PEDSSRSFIFPGGSLSCVGEKFAKYLKFLVELTOHCNMLILSNPPTKGTPTYPV 478
Db 419 PEDSSRSFIFPGGSLSCVGEKFAKYLKFLVELTOHCNMLILSNPPTKGTPTYPV 478
QY 480 DNLPTKFTSY 489
Db 479 DNLPAKIFG 488

RESULT 2
Q9R1F4 PRELIMINARY; PRT: 497 AA.
ID 09R1F4
AC 09R1F4;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DE 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE CYTOCHROME P450 RETINOIC ACID METABOLIZING ENZYME P450RA.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TOOTH;
RX PubMed=11063033;
RA Paine C.T., Paine M.L., Sneed M.L.;
RT "Identification of tufelulin- and amelogenin-interacting proteins using
the yeast two-hybrid system";
RL Connect. Tissue Res. 38:257-267(1998).
CC -1- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) = ROH +
OXIDIZED FLAVOPROTEIN + H(2)O (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM (BY
SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC EMBL: AF115769; AAD17217.1; -.
DR InterPro: IPR001128; -.
DR Pfam: PF00067; P450; 1.
DR PRINTS: PR00385; P450.
DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
KW Electron transport; Endoplasmic reticulum; Heme; Membrane; Microsome;
Monooxygenase; Oxidoreductase.
SQ SEQUENCE 497 AA; 56191 MW; 67CB167A997842C1 CRC64;

```

Query Match 67.8%; Score 1738; DB 11; Length 497;
Best Local Similarity 67.9%; Pred. No. 2.4e-107;

```

Matches 337; Conservative 60; Mismatches 89; Indels 10; Gaps 4;
QY 1 MGLYTLMTVECTIVLPVLLFLAAYKIMEMILIRVDDPCRSPLPFGMGLPFIETIOL 60
Db 1 MGLPALATLCTFVLLFLAALKMLDLYCVSSRDSICALPLPFGMGLPFIETIOL 60
QY 61 ILORRKLPMKROKGYCYTHLFGNPTVVMGADNVROILLGHEKLVSVOMPASVETIL 120
Db 61 VLORRKLPMKRRKGYCYTHLFGNPTVVMGADNVROILLGHEKLVSVOMPASVETIL 120
QY 121 GSDPLSNVHGVQHNKKKAIKIMAFSRDALHPIVIOEVSALIOEMLOKDCVLYPEM 176
Db 121 GSGCLSNLHNGOHRKKKIVIMQAFSRDALHPIVIOEVSALIOEMLOKDCVLYPEM 177
QY 177 YPEMKIEMIRAMRILGEPBOQITD-EQELVEAFEMIKNLSLPIDVPFSGLYRGL 234
Db 177 YPEMKIEMIRAMRILGEPBOQITD-EQELVEAFEMIKNLSLPIDVPFSGLYRGL 237
QY 235 RANFTHSKIEENIRKRIODDNEENOKYKDALQILLIENSRRSDPEFSLOAKKEATL 293
Db 235 RANFTHSKIEENIRKRIODDNEENOKYKDALQILLIENSRRSDPEFSLOAKKEATL 297
QY 294 LFGCHETASTATSLVWFLDNLNTEVVOKREVEQVEKMGMTTPEGKLSMELLDOILKYTG 353
Db 294 LFGCHETASTATSLVWFLDNLNTEVVOKREVEQVEKMGMTTPEGKLSMELLDOILKYTG 357
QY 354 CVTKETLRINPVPVGGFRALKTLELNGYQIPKGMNVTISICDTHVADVPKKEEOPREF 413
Db 354 CVTKETLRINPVPVGGFRALKTLELNGYQIPKGMNVTISICDTHVADVPKKEEOPREF 417
QY 414 RFMSKGLDESGRENYTFPGGSRMCVGEKFAKYLKFLVELTOHCNMLILSNPPTKGTPTYPV 473
Db 414 RFMSKGLDESGRENYTFPGGSRMCVGEKFAKYLKFLVELTOHCNMLILSNPPTKGTPTYPV 477
QY 474 PTIYPVDNLPTKFTSY 489
Db 478 PTIYPVDNLPTKFTSY 493

RESULT 3
Q93323 PRELIMINARY; PRT: 492 AA.
ID 093323
AC 093323;
DT 01-NOV-1998 (Tremblrel. 08, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DE 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE CYTOCHROME P450 26 (BC 1.14.-.-) (RETINOIC ACID-METABOLIZING
CYTOCHROME) (P450RAI) (RETINOIC ACID 4-HYDROXYLASE).
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RA Chen J.L., Grunz H., Panitz F., Pieler T., Holleemann T.;
RT Submitted (Apr-1998) to the EMBL/Genbank/DBP databases.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC EMBL: AF057566; AAC25158.1; -.
DR InterPro: IPR001128; -.
DR Pfam: PF00067; P450; 1.
DR PRINTS: PR00385; P450.
DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
KW Heme; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 492 AA; 55459 MW; D1D4BB7651BF2D3E CRC64;

```

Query Match 64.2%; Score 1646.5; DB 13; Length 492;
Best Local Similarity 63.2%; Pred. No. 2.8e-101;
Matches 312; Conservative 84; Mismatches 87; Indels 11; Gaps 4;

```
Db 1 MDYTLTSLAICTLALPLLLLLTAAKIMEVYCLRRKADACANPLPCTMGLPFEGGTLOM 60
Qy 61 ILORRFLRMKROKCYCTKTHLFGNPTVRMGADNVROILLGCHKVSVQMPASVPTIL 120
Db 61 VLORRFLQKRSQYGYKITYHFLGSPVTVTAENVROILLGCHKVSVHWPASVPTIL 120
Qy 121 GSDTLNVHGVQHKRKAIMRAFSDALEHYIPVIOEYKSAIOEWLOKDCSVLYPEM 180
Db 121 GAGCLSLNHDHEKRYTKVLAQAFSREALANYVOMEVEEYRCSYVNLQSGPCVLYYPAI 180
Qy 181 KLMFRIAMRILLGFEPQIKTDEQ-LVFAFEEMIKNLSPIDVPFSGLYRGLRARNF 239
Db 181 KRMFRIAMRILLGCDPQRDRDEQETLLEAFEMSKNLSPLIDVPFSGLYRGLRARNL 240
Qy 240 IHSKIEENIRKRIODDNEQKYKDALQILLIENSRRSDPEFSIQAMKEATELLFGGHE 299
Db 241 IHAQIEENIRKRLQ--REPDEHCKDALQILLIYSSRRNGEPINLOALKESTELLEFGGHG 297
Qy 300 TTAATASILVFLGLNTEVVOKYREVEQEKYEMGMT---PGKLSMELLQDKYTCGV 355
Db 298 TTASASTSLTFLAHKDYLEKVRKKELETO---GLLSTRPEKKELESLIVLQOLKYSV 354
Qy 356 IKETLRINPVPVGGFRVALKTFELNGYQIPKGMNVIYSCDTHVDVADVPKKEEPOPER 415
Db 355 IKETLRISPVPAGGFRVALKTFVNLGYQIPKGMNVIYSIADTHEADLPPTDKFNDFRE 414
Qy 416 MSKGLDEDSRFNYIPFGGSRMCYKKEFAKYLKIFLVELTQHCNWLISNGPPTMKTGPT 475
Db 415 LTPRLRDSRRGFIPIFGGVCVRCIGKEPAKILKVFVELCRNCDWELNGLSPAMTSP 474
Qy 476 IYVDNLPTKFTSY 489
Db 475 ICPVDNLPAKFKPF 488

RESULT 4
Q9NR63 PRELIMINARY: PRT: 512 AA.
ID Q9NR63
AC 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE CYTOCHROME P450 RETINOID METABOLIZING PROTEIN P450RA1-2.
GN CYP26B1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE-20300913; PubMed-10823918;
RA White J.A., Ramshaw H., Taimi M., Stangle W., Zhang A., Everingham S.,
RA Creighton S., Tam S.-P., Jones G., Petkovich M.;
RT Identification of the human cytochrome P450, P450RA1-2, which is
RT predominantly expressed in the adult cerebellum and is responsible for
RT all-trans-retinoic acid metabolism."
RT Proc. Natl. Acad. Sci. U.S.A. 97:6403-6408(2000).
CC -1- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) = ROH +
CC OXIDIZED FLAVOPROTEIN + H(2)O (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC EMBL: AF252297; AAF76003.1; -.
DR InterPro: IPR001128; -.
DR Pfam: PF00067; P450; 2.
DR PRINTS: PR00385; P450.
DR PROSITE: PS00086; CYTOCHROME P450; UNKNOWN.1.
KW Electron transport; Endoplasmic reticulum; Heme; Membrane; Microsome;
KW Monooxygenase; Oxidoreductase.
SQ SEQUENCE 512 AA; 57512 MW; A06D1D9944E6726F CRC64;
```

Query Match

40.4%; Score 1034.5; DB 4; Length 512;

```
Best Local Similarity 42.3%; Pred. No. 1e-60;
Matches 209; Conservative 100; Mismatches 166; Indels 17; Gaps 8;

Qy 1 MGLYTLMTVECTIVLPLVLLAAYKEMLMIRVDPNCRSPRLPPGTMGLPFIEETLOL 60
Db 6 LDVLSATLALACIVSVTLILAVSOOLMRMAATRDCKSLPIPGKSGFGLIETGHW 65
Qy 61 ILORRFLRMKROKCYCTKTHLFGNPTVRMGADNVROILLGCHKVSVQMPASVPTIL 120
Db 66 LLOGSGFQSSRRKRYGVNTRKTHLGRPLRVYGAENVKRIIMGENHIVSTEMPSTRMIL 125
Qy 121 GSDTLNVHGVQHKRKAIMRAFSDALEHYIPVIOEYKSAIOEWLOKDCSVLYPEM 180
Db 126 GPRTVNSIGDIRKNRKVFSLFSEHALESYLPKIQVLVOTLIRAMSSHPREAINVYQEA 185
Qy 181 KLMFRIAMRILLGF-PEQIKTDEBELVEAFPEEMIKNLSPLIDVPFSGLYRGLRARNF 239
Db 186 OKLTFRMAIRVLLGFSPER--DLGHLEEVYQOFVNDVSLPVDLPFSGYRGIQARQI 242
Qy 240 IHSKIEENIRKRIODDNEQKYKDALQILLIENSRRSDPEFSIQAMKEATELLFGGHE 299
Db 243 LQKGLKAIKEKIQCTQGRD---YLDALDLIESSKEHGEKEMTQKOTLELIPAAVA 299
Qy 300 TTAATASILVFLGLNTEVVOKYREVEQEKYEMGMT---TPKGL-SMELLQDKYTCG 354
Db 300 TTASASTSLIMQLKHPYLEKLRDELRAH---GIHSGGCPCEGTLRLDTLSGLRYLDC 356
Qy 355 VIKETLRINPVPVGGFRVALKTFELNGYQIPKGMNVIYSCDTHVDVADVPKKEEPOPER 414
Db 357 VIKEMRLFTPIISGTYVLTQTELDGFOIPKGMNVIYSTRDHDAVPKDVNVDPDR 416
Qy 415 FMSKGLD-GSRFNYIPFGGSRMCYKKEFAKYLKIFLVELTQHCNWLISNGP-PTMKT 472
Db 417 FSOARSEDKGRFHYLPFGGVRCTCKHLAKLFLKVALELASTSFELATRTPTPRTIL 476
Qy 473 GPRTVVDNLPTKF 486
Db 477 VPLVLPVDGLSVKF 490

RESULT 5
Q9NP41 PRELIMINARY: PRT: 525 AA.
ID Q9NP41
AC 09NP41;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE WDGSCH.NH0493L16.1 PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE-99063792; PubMed-9847074;
RA Sulston J.E., Waterston R.;
RT "The sequence of Homo sapiens BAC clone RP11-493L16."
RT "Toward a complete human genome sequence."
RL Genome Res. 8:1097-1108(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA Bourne S., Bauer C., Pape K., Jones T.;
RT "The sequence of Homo sapiens BAC clone RP11-493L16."
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Waterston R.;
RT Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) = ROH +
CC OXIDIZED FLAVOPROTEIN + H(2)O (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC EMBL: AC007002; AAF65576.1; -.
```

DR InterPro: IPR001128; -
DR Pfam: PF00067; P450; 1
DR PRINTS: PR00385; P450.
DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
KW Electron transport; Endoplasmic reticulum; Heme; Membrane;
KW Monooxygenase; Oxidoreductase;
KW SEQUENCE 525 AA; 59124 MW; 71DA7B6752A60315 CRC64;

Query Match	39.9%;	Score 1022;	DB 4;	Length 525;
Best Local Similarity	41.2%;	Pred. No. 7.1e-60;		
Matches 209;	Conservative 101;	Mismatches 167;	Indels 30;	Gaps 9;

QY	1	MGVLINMTEFCTIVLPLVPLLEFLAIVKIMEMIMIRVPDNCRSPRLPQTMOLJPRJGZHL	60
Db	6	LDVLSATLATAACLVSVTLILAAVSQOLMQRMAATRDSCKRLPIKQSMQFPIIGENGM	65
QY	61	ILOR-----RKFLRMKQKVCGLYKTHLGNPTVMADNRQILGSHKL	107
Db	66	LLQKLTLETRHVMPLRQSGSGFOSSRREKIGANFKHLHGRPLIRYTGAEVNRKILMGSHHL	125
QY	108	VSVOGPAVSRTIIGSDPLSNVGHVOHKKKALIMAFSDALEHYIPVIOEQEKSAIQEW	167
Db	126	VSTEMPRSTRMLLPNTVNSIGIHNNKRVKESKISHLSHELSYKRLQVLODTLRAM	185
QY	168	LQKSCVLVYVEMKMLFRIAMRLLEFE--PEQIKTDEOEIVAEFEEMINLSPLDVP	226
Db	186	SSHDEAIVVYOEAOKLTFEPMARIVALLFSFISPE--DLGHLEFYVOQVNVFSLPDLR	242
QY	227	FSQYRGRLRANFNHSHKEENIRKIDODDENQOKRDLLOLLENSRSDPEFSIOAM	286
Db	243	FSQYRGRLQANQIIOKLEKAIIRKLCQTOGKD---YLDMLDLIESKHNKHEMIMQEL	299
QY	287	KEATLELFGSHETTASTATSLVYVFLGNTPEYVOKVAEEVQEKYEMMY-----TPKGG-L	341
Db	300	KDGTLELLEFAAYATTASASTSLIMOLKLHPVLEKTLDELRAH--GILSHGGPCPGTLL	366
QY	342	SMELLDOUKYACVKEFLRINRPVPGSFYVALTFEELNGVOYIPKGMNVYSIDTIDHVA	401
Db	357	RLDLTSLGRYLDVCVKEWMRLFTFISGYSRVLTOTFELDGOIPIKQSWMYSIDTIDHVA	416
QY	402	DVFPKKEEFOFERMKSQKLE-GSRPNYIPPIGGGSRMCVCKEFPKVLKLFIVELTHGN	460
Db	417	PVFKNVYVNEPDDRQSQAESDKDGRFHLPFGGVSVCIGLHKLAKLPLKLVANVELASTR	476
QY	461	WILSNGR-PIYKGTPTIYPVNDLPTKF	486
Db	477	FELATRTFETPLVPLHPVDGLSVK	503

RESULT	6		
059990			
ID	059990;	PRELIMINARY;	PRT; 444 AA.
AC	01-NOV-1996 (TREMBLrel. 01, Created)		
DT	01-NOV-1996 (TREMBLrel. 01, Last sequence update)		
DT	01-MAR-2001 (TREMBLrel. 16, Last annotation update)		
DE	PUTATIVE CYTOCHROME P450 120.		
GN	CYP120 OR CYP OR SLR0574.		
OS	Synechocystis sp. (strain PCC 6803).		
OC	Bacteria; Cyanobacteria; Chroococcales; Synechocystis.		
OX	NOBL_TaxID=1148;		
RP	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-PCC6803;		
RA	Tabata S.;		
RL	submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=96127529; PubMed=8590279;		
RA	Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,		
RT	Sugita M., Tabata S.;		
	Sequence analysis of the genome of the unicellular cyanobacterium		

RT Synechocystis sp. strain PCC6803 I. Sequence features in the 1 Mb
RT region from map positions 64% to 92% of the genome.";
RL DNA Res. 2:153-166(1995).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; Pubmed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hikosawa M., Sugliua M., Sasamoto S., Kinura T.,
RA Hosouchi T., Matsuo A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA Shampo S., Takeuchi C., Wada T., Watanabe A., Yamada M.,
RA Yabuta S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL: D64003; BAA10496.1; -;
DR InterPro: IPR001128; -;
DR Pfam: PF00067; P450; 1;
DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
DR Hypothetical protein; Heme; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 444 AA; 50578 MW; 8f62a5e03b54bdc CRC64;

Query Match	26.9%	Score	689.5	DB	2	Length	444
Best Local Similarity	34.5%	Pred. No.	6e-38				
Matches	156	Conservative	99	Mismatches	170	Indels	27
						Gaps	10

[illegible]

RESULT	7	
09LIVY7		
ID	09LIVY7	PRELIMINARY; PRT; 477 AA.
AC	09LIVY7	
DT	01-OCT-2000 (TREMBLrel, 15, Created)	
DT	01-OCT-2000 (TREMBLrel, 15, last sequence update)	
DT	01-MAR-2001 (TREMBLrel, 16, last annotation update)	
DE	CRYOCHROME P450-LIKE.	
OS	Arabidopsis thaliana (Mouse-ear cress).	
OC	Eukaryota, Viridiplantae: Embryophyta, Tracheophyta: Spermatophyta:	
OC	Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II:	
OC	Brassicales; Brassicaceae; Arabidopsis.	

```

OX NCBI_TaxId=3703;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA.
RX MEDLINE=20181125; PubMed=10718197;
RA Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Kotani H.,
RA Tabata S.,
RT "Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence
RT features of the regions of 3,076,755 bp covered by sixty P1 and TAC
RT clones."
RL DNA Res. 7:31-63(2000).
CC -1- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) = ROH +
CC -1- OXIDIZED FLAVOPROTEIN + H(2)O (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL, AB018112; BA96885.1; -.
DR InterPro, IPR001128; -.
DR Pfam, PF00067; P450; 1.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_1.
KW Electron transport, Endoplasmic reticulum, Heme, Membrane, Microsome,
KW Monooxygenase, Oxidoreductase.
SQ SEQUENCE 477 AA; 54850 MW; 30AB4B2C4970D405 CRC64;

```

Query Match	23.48;	Score 598.5;	DB 10;	Length 477;
Best Local Similarity	31.08;	Pred. No. 7.1e-32;		
Matches 155; Conservative	94;	Mismatches 182;	Indels 69;	Gaps 18

```

QY 14 IYVPLVFLAAKLVEMIMIRVOPNCRSP-LPPGWMGLPFGEJLOLILQR-----KF 67
   1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3 MAMILIFLSTLSLSDLLLRKHLSHSYNPLPQNGTGLPLGSEFSLSAGROGHEKF 62
   1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 68 L--RMR--QKGCITCYKTHLPGNFTYVWGVADNROILGHEKLVSVOPASVFTILGSD 123
   1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 63 ITRVRFSSSSSSCFKTHLGEPTAVVGTASGNKFLFTNENKLVSWMPDSVNRKIPSS 122
   1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 124 TLSNVGVOKHKKKKAIABRSP-DALVHTPIVIOOEVSALO-EMLOKDSCLVYPEMK 181
   1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 123 MQR--SKERAKLRLMLSOPEALRKYVGMELIARHETEMANQO-VIYFPLTK 179
   1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 182 KLMFRIAMKILGFE-PEQITDQELVEAFEEIMKINLPSLIDVPFSGLYRGLRANFT 240
   1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 180 KFFFSIACRSFLSMEDPARVR---OLEOEPNTVAAGISFISIDLPCTRFRNAIKASRL 235
   1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 241 HSKIEENIKRLODDONEEOKYKDALO-----LLIENSRSRSEPSLOAMEKATEL 293
   1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 236 RKVSALVQR-----KEELKAKKALEEHDILSHLMIMIGETKED-----LADKITGL 284
   1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 294 LFGGHETASTATSLVMEFLGNTVEVOKREE---VOEREMGMYTPGKGLSMELLDOU 349
   1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 285 LIGHGFAIVCTEFVNYLAEFPVYORVLOEOKELIKKEK-----EGLRMEDEIKM 338
   1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 350 KYNGCYIKETLINPPVPGGRVALKFTLNGOIPKGNNAVYSICDTHDADVDFPKKE 409
   1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 339 RYSMWACCVMTIYVPLSGTIRFALIDFESKGYIPKGNKLWTSATAHNMPDYFEPEP 398
   1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 410 FODERPMASKLEDGSRFNVIIPFGGSMCVGKFAFVLLKILVPLVLOHGMWLLSNPPT 469
   1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 399 FEENRFEGSGPKP---YTVYVFGGGRPMCGKREYARLELILPMHNLVNRFW----- 447
   1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 470 MKGTPIY-----VDULP 483
   1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 448 ----EKVFPENKIVYDULP 463
   1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT	8	
Q9SJH2		
ID	Q9SJH2	PRELIMINARY;
AC	Q9SJH2;	PRT; 485 AA.
DT	01-MAY-2000	(TRENDAurel_13, Created)
DT	01-MAY-2000	(TRENDAurel_13, Last sequence update)

DT 01-MAR-2001 (Tremblay et al. 1996, last annotation update)
DE PUTATIVE CYTOCHROME P450.
GN AT2G42850.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;

RP SEQUENCE FROM N.A.
RC STRAIN-CV. COLUMBIA;
RX MEDLINE-20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
RA Cronin L.A., Shen M., Yankson S.E., Unyeam L., Tallon L.J., Gill J.E.,
RA Adams W.D., Carriera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA Copenhagen G.P., Preuss D., Nierman W.C., White O., Eissen J.A.,
RA Salzberg S.L., Fraser C.M., Venter J.C.;
RT "Sequence and analysis of chromosome II of *Arabidopsis thaliana*,"
RL Nature 402:761-768(1999).
CC -1- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) -> ROH +
CC OXIDIZED FLAVOPROTEIN + H(2)O (BY SIMILARITY).
CC
CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM (BY
CC SIMILARITY).
CC
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC EMBL, AC006931; AAD21724.1; -.
DR InterPro: IPR001128; -.
DR Pfam: PF00067; P450; 2.
DR PRINTS: PS00385; P450.
DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
KW Electron transport; Endoplasmic reticulum; Heme; Membrane; Mitosome;
KW Monooxygenase; Oxidoreductase;
KW Sequence; 485 AA; 55405 MW; 08B0B16474620F82 CRC64;
KW

Query Match	20.8%;	Score 532.5;	DB 10;	Length 485;
Best Local Similarity	28.2%;	Pred. No. 1.7e+27;		
Matches 143;	Conservative 101;	Mismatches 214;	Indels 49;	Gaps 15;

OY	4	YTLMAVFLC--TIVYRVLVFLAAVKAEMIMIRVBPNCSPILPBTGMPLPGLTQOLI	61
Db	7	FVLSWFLCIATITISSTLFEFR--KNNHRTFKIKO--KKKLLPBGKMLPWIGETMDFY	62
OY	62	LOHRK-----FLPMKROKXGICITKTLFNGPPYRVWAGADNNOILIGENKLVSNOWAS	115
Db	63	KAKSKNRVDEVDVNRILIKHGNIKFTRIMSPPTIYNAGAEARLLISNFSIYVSNWSS	122
OY	116	VRTILCSPLTSLNVGVOYKRNKKKAIIMAFSFDALERTIPIYQOEKSAIQ--EMLQDCSV	174
Db	123	SVOLMNMCMIMAKOEKRRHVRGLVANSLSLTIGLSTLPLCDYKAFHHEWRKKEE-I	181
OY	175	LVYPEKKKIMFRIAMRILGLGEPRQIKTBOELVEAFEEIKNLSLPIDVPSGLYRGL	234
Db	182	SLYRSKAKVLTFTVFECLYG-----IKV-EIGMLEVEFERYLEGEVFLAPYEPFCSFARAK	235
OY	235	RAKNPFIHSIEINIRKKIITODDENEBOEKYDAL-QLLIENSRSDPEFSLOAKRAATE-	292
Db	236	KALHELETFLVGVKREKREMEKSGAEKRPNTLFSRLVE-----ELIKVITTEE	288
OY	293	-----LLFGCHETATSTATSLVMEFLGLNTEYVOK--VREYOEKVEKMTTPRKGISM	343
Db	285	EYVDNNVLLVFAAHDTSTAMSMTEFKMLAOPHCRODTLLOEHAQIKANG--EBEYLTIV	341
OY	344	ELLDDOLKUYGCAIKETLRINPVPBGSGRVALLKFEELNGYOIRKGNVNVYISICDPIADVAD	403
Db	342	EDVKKKKKYUQVYRETRMLSPPIGSGPRKAAVADIDGQYTIPIKGMILMTITGYHNPDEI	401
OY	404	EPKKEEPOEERMSGLGDSGRNYTIRPPGSGSMYCGKFEAVLLKIFLVELDHOHML	466
Db	402	FQDPMSFDPTNR-DKPIQ---ATYTLRPPGSGPRLCAGHOLAKISIVENHNYVYGFBSL	457
OY	464	SNGPPTMKTGPTIYVDNLPRKFTSYV	490

Db 458 YYPDETISMDPLPFPSPLOMPKISPKV 484

RESULT 9
ID 09FH76 PRELIMINARY; PRT: 463 AA.
AC 09FH76;
DT 01-MAR-2001 (TREMblrel. 16, Created)
DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)
DE CYTOCHROME P450.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RX MEDLINE=20181125; PubMed=10718197;
RA Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Kotani H.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence
RT features of the regions of 3,076,755 bp covered by sixty pl and TAC
RT clones";
RL DNA Res. 7:31-63(2000).
DR EMBL; AB020744; BAB10255.1;
SQ SEQUENCE 463 AA; 52366 MW; CCD17293F53F812 CRC64;

Query Match 20.4%; Score 523; DB 10; Length 463;
Best Local Similarity 29.4%; Pred. No. 6,8e-27;
Matches 145; Conservative 93; Mismatches 203; Indels 52; Gaps 15;

QY 2 GLYTIM--VFLECTVLPVLFIAAVKLMEMLRVDPNCRSLPFGTGLPFGFTL 58
DB 5 GLEFLSAAFLC-----LIRFLAGV-----RRSSSTKLPLPGTGYVYGTEF 50
QY 59 QLIID-RKRFPMRKOKGYKTHLFGNPTVRVAGADVQOILGELKILSVQMPASVR 117
DB 51 QLSQDPNVPFAKORRGVFKTHVLGCPVMAISSPEAKFVLVTKSHLKKPFPASK 110
QY 118 TILGSDTLNVHGVQHNKKKAIKRAFSRDLNHYIPVIOGVKSAIOEWLOKDSVLT-V 176
DB 111 RMLGQALFEFHOGYHSLRKVLRAFMPIRNVVPIESIAQESLNSW--DGTQINT 167
QY 177 YPEMKKLEFRIAMRLLGLFEPEQIKTDEQELVEAFEEIKMLFSLPIDVPSGLYGRRA 236
DB 168 YQEMKTYTFNVALISILKDEYYRED--LKRCYILLEKGYNSMPILPGLTFHAKAKA 224
QY 237 RNFHSKLEINIRKKIQQDDNENQKYDALQLLIEN-SRSDEPFSLOAKKATLLEF 295
DB 225 RKEL-AQILANILSK-----RRONPSSHDLGSEMEDKAGLTDQIA-----DNIGVIF 274
QY 296 GGHETTATATSLVFLGTLNTEVQVQVREYV---QEVLEMGMTTPGKSMELLDQIKY 351
DB 275 AARDTTASVLTWILKYLDNPTVLEAVTEEMAIKDKKE-----GSLTWEDTKKMP 328
QY 352 TGCVIFETLRINPVPYGGFVALKTFELNGYQIPKGNVIVSICDTHDADVFPNKEEFQ 411
DB 329 TYRVIOETLRATILISFPFRAVEDEVEGYLIPKWKVLDLPLFRNIHNDIFSDPGFD 388
QY 412 PERFMSKLEEGDSRN-TIPFGGSRMCVKEPAKVLKILVELTQHCHNILLSNGPTM 470
DB 389 PSRE-----EVAAPKPNTPMGSGIHSCPNELAKLEISVLIIHLTTKYRMSIVGPSDGI 443
QY 471 KTGPPIYVDNLP 483
DB 444 QYGPALPONGLP 456

RESULT 10

Q9LKH7
ID Q9LKH7 PRELIMINARY; PRT: 474 AA.
AC Q9LKH7;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)
DE CYTOCHROME P450.
GN CYP7A.
OS Phaseolus aureus (Mung bean) (Vigna radiata).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
OC Fabales; Fabaceae; Papilionoideae; Vigna.
OX NCBI_TaxID=3916;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. 2937;
RA Yang M.T., Chen Y.M.;
RT "Cloning and sequencing of a Vigna radiata cDNA encoding cytochrome
RT P450.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) = ROH +
CC OXIDIZED FLAVOPROTEIN + H(2)O (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL; AF279252; AAF89209.1; -.
DR InterPro; IPR001128; -.
DR Pfam; PF00067; P450; 2.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_1.
KW Electron transport; Endoplasmic reticulum; Heme; Membrane; Microsome;
KW Monooxygenase; Oxidoreductase.
SQ SEQUENCE 474 AA; 54037 MW; A2542A809C5BAC6D CRC64;

Query Match 20.1%; Score 515.5; DB 10; Length 474;
Best Local Similarity 29.4%; Pred. No. 2.2e-26;
Matches 141; Conservative 99; Mismatches 209; Indels 31; Gaps 15;

QY 14 IVPLVLFIAAVKLMEMLRVDPNCRSLPFGTGLPFGFTLQILQRRK-----FL 68
DB 2 VSLPTLLILFLAASA-AAIFLHRAFSRRKRLPFGSGILPFGFTLQILAYSSNDEP 60
QY 69 RMRQKGYCYKTHLFGNPTVRVAGADVQOILGELKILSVQMPASVRLTIGSDLSNV 128
DB 61 DERVRRGSGIFMTHVGEPTVESADELNRFTILQNGKLLDSYPSISNLLGKSLILM 120
QY 129 HGVOHNKKKAIKRAFSRDLN-EHYIPVIOGVKSAIOEWLOKDSVLYPEMKKIMPR 186
DB 121 KGLHKK-RHMSLSTFNSANSITKHLLHIDRLGLNLTWSDR--VILMOAKKITFE 176
QY 187 IANRILGFPEQIKTDEQELVEAFEEIKMLFSLPIDVPSGLY-RGLARNFHSKITE 245
DB 177 LVYQILMSFDPDEM--TESIRKEYVLVIEGFTPLPL-FSTYVRAIKARVAAEALT 232
QY 246 ENIRKKIQQDDNENQKYDALQLLIENSRSDPESLOAKKATLLEFGHETTASTA 305
DB 233 LVVRQR-RETYNQGKESKSMGALL---ASGDHSDQIYDFELALLVAGIETTSTIM 287
QY 306 TSLVMEGLNTEVQVQVREYQVEMGYTPGKISMELLDQIKYTCVIFETLRINP 365
DB 288 TLAVKLEETPLLAOLKEH-HDQIR-ARSDPGAPLEMTDYKSMVETQVAVNEFLRVANI 345
QY 366 VPGGFVALKTFELNGYQIPKGNVIVSICDTHDADVFPNKEEFQPERFMSKLEEGDSR 425
DB 346 IGGIFRATTDIDIKGYTTPPKWKVAFASRAVAILNDEYTKDARTFPMWQSNSSSEANP 405
QY 426 FN-YIPFGGSRMCVKEPAKVLKILVELTQHCHNILLSNGP-----PTMKTGPPIYV 479
DB 406 ANYTTPFGGPRICPGYELARVLSVFLHRYTRFSWVAPEEDKLVFFPTTTRQKR-VPI 464

RESULT 11

081077 PRELIMINARY; PRT; 482 AA.
 ID 081077
 AC 081077;
 DT 01-NOV-1998 (Tremblrel. 08, Created)
 DT 01-NOV-1998 (Tremblrel. 08, last sequence update)
 DT 01-MAR-2001 (Tremblrel. 16, last annotation update)
 DE PUTATIVE CYTOCHROME P450.
 GN T914.17.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 OC Brassicales; Brassicaceae; Arabidopsis.
 NCBI_TaxID=3702;
 RX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA Rounsley S.D., Lin X., Kaul S., Shea T.P., Fujii C.Y., Mason T.M.,
 RA Shen M., Rensing C.M., Fraser C.M., Somerville C.R., Venter J.C.,
 RT "Arabidopsis thaliana chromosome II BAC T914 genomic sequence."
 RL Submitted (Aug-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 DR EMBL: AC005315; AAC3335.1; -
 DR InterPro: IPR001128; -
 DR Pfam: PF00067; P450; 1.
 DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
 KW Heme; Monooxygenase; Oxidoreductase.
 KW SEQUENCE 482 AA; 55175 MW; AB6A07AB2778D3B CRC64;

Query Match 19.6%; Score 503; DB 10; Length 482;
 Best Local Similarity 28.9%; Pred. No. 1.5e-25;
 Matches 140; Conservative 92; Mismatches 224; Indels 28; Gaps 11;
 QY 7 MVTFLCTVPLVLLFLAVALKLMEMIRVDPNCRSPPLPGTGLPIGTLQILQ-R 65
 DB 22 LITLTIIVVVVLLF---KWLHMKRQRL-----RLPSCMGLPIGTGLRXYTENPN 71
 QY 66 KFLMRKRKQYGCYKTHLFGNPTVRYMGADNVRQIILGKHLYSVQMPASVRTL 125
 DB 72 SFATRQNKYGDIFKTHLGGPCVWISSPEARVLSKAHLKPTYPSPKERMIGPAL 131
 QY 126 SNVGVQHNKKKAIMRAFSDALEHYTPVIOEYKSAIOEML-QKDSCLVYPMKKLM 184
 DB 132 FEHGPYISTKLRLVQSSFMPSALRPYSHLELVQLTSSWTSOKSINTLEY--MKRYA 189
 QY 185 FRIARILLGFEPIQKTDEQELVAEEMTKNLSPLIDVPSGLYRGLEARNFHSKI 244
 DB 190 FQVAIMSAFGCKEPTTIDVIKLL--YQRLRGYNSMPLDPLGTLFHKSMKAR----IEL 243
 QY 245 EENIFKKIQDDNENEOKYKDALQILLNSRRSDPEPSLOAMKEAATLFLGHEHTTAST 304
 DB 244 SEELKKV--EKRRNGREEGGLLVLLGAKQKNGLSDSQIADNITGVITAAIDTTASV 302
 QY 305 ATSLVMTGLNTEVQKREVEQKEMGMYTPGKGLSMELLQIKYTCVYIKETLRINP 364
 DB 303 LTMILKYLHDPNILLQEVSR--QFSIRQIKKENRISMEPTRKMPLTRVYIQETLRAS 361
 QY 365 PVPGGFRVALKTFELNGYQIRKGNVITYISCTDHDVADVPFNKEFQPERMSKLEGS 424
 DB 362 VLSFTRFREVQDVEVDGLIPGMKVLPFLFRIRHSSSEFPDPEKFSR-----EVAP 416
 QY 425 R-FNTIPFGGGRMCVGEFPAVKLLKFLVELTQHCNMLISNGPTKGTPIYVDNLP 483
 DB 417 KRYTTPRGNGVHSCPGSELAKEMLILHLHLLTSFRKEVIGDEGIDGYGPPVYKKGLP 476
 QY 484 TKFT 487
 DB 477 IRVT 480

RESULT 12
 ID 064989 PRELIMINARY; PRT; 513 AA.

AC 064989;
 DT 01-AUG-1998 (Tremblrel. 07, Created)
 DT 01-AUG-1998 (Tremblrel. 07, last sequence update)
 DT 01-MAR-2001 (Tremblrel. 16, last annotation update)
 DE STEROID 22-ALPHA-HYDROXYLASE.
 GN DMF4.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 OC Brassicales; Brassicaceae; Arabidopsis.
 NCBI_TaxID=3702;
 RX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=WS-2;
 RA Choe S., Dilkes B.P., Fujioka S., Takatsuto S., Sakurai A.,
 RA Feldmann K.A.,
 RL Plant Cell 10:0-0(1998).
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 DR EMBL: AF044216; AAC05093.1; -
 DR InterPro: IPR001128; -
 DR Pfam: PF00067; P450; 1.
 DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
 KW Heme; Monooxygenase; Oxidoreductase.
 KW SEQUENCE 513 AA; 58867 MW; B1639BDD9A4DA6F3 CRC64;

Query Match 19.6%; Score 503; DB 10; Length 513;
 Best Local Similarity 27.6%; Pred. No. 1.7e-25;
 Matches 144; Conservative 95; Mismatches 216; Indels 66; Gaps 14;
 QY 4 YTLVMTFLCTVPLVLLFLAVALKLMEMIRVDPNCRSPPLPGTGLPIGTLQIL-- 61
 DB 7 HTLLPILLPLSLSLFL-----ILKRNKRTRNRLPGRKGMFLGTLGTYLNP 58
 QY 62 --LQRRFLMRKQYGCYKTHLFGNPTVRYMGADNVRQIILGKHLYSVQMPASV 118
 DB 59 YTAATLGDGFQOHVSKYGGKIRNSLFGSEPTIVSADAGLNFILQNGRFLFECSSYPSISG 118
 QY 119 ILGSDTLNHNHGVQHNKKKAIMRAFSDALEH-----YIPVIOEYKSAIOEMLQKDS 172
 DB 119 ILGKWSMLVLVGMHNRD---MRSISLNFLSHARLTLILKVEKHTLFLVDSW-QONS 172
 QY 173 CVLYVPMKKLMFRIAMRILLGFEPIQKTDEQELVAEEMTKNLSPLIDVPSGLYR 232
 DB 173 IFSQDPAKKEFTFNLMAKHIMSMDPGEET--EQLKKEYTFPKGVASAPLNPGLAYHR 230
 QY 233 GLRAR---NFHSKIEENIRKKIQQDDNENEOKYKDALQIL-----IENSRRSD- 279
 DB 231 ALQSRATILFKTEKMEERKIDIKKEQDEEEVKTDEEAMSKSDHYRQRQTDLDLGMV 290
 QY 280 ----PFSLOAMKEAATLFLGHEHTTASTATSLVMTGLNTEVQKRE--VOEKYE 331
 DB 291 LKHSNLTSEQLDLILSLRAGHETSSVAIALAIFLQACPKRAVELREHLEIAAKKE 350
 QY 332 MGYTPGKGLSMELLQIKYTCVYIKETLRINPVPGGFRVALKTFELNGYQIRKGNV 391
 DB 351 LG-----ESELNMDQYKMDFTQCVINETRLGNVVRILHAKKALDKVYKKYDIPSGMKVYL 406
 QY 392 YSICDTHDVAQVFNKEFQPERM-----SKGLEDSRF--NYIPFGGGRMCVGRK 442
 DB 407 PVISAHLDNSRDQPNILFPMRWQOQONNAGSSGSSFTWGNQNPFGGPRGLCAGSE 466
 QY 443 FAKVILKIFVELTQHCNMLISNGPTKGTPIYVDNLP 483
 DB 467 LAKLEMAVFIHLVLKFNWELAEDDQ-----PFAFPVDP 502

RESULT 13
 ID 09SC09 PRELIMINARY; PRT; 513 AA.
 AC 09SC09;
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, last sequence update)

DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
 DE STERIOD 22-ALPHA-HYDROXYLASE (DWF4).
 GN 1345.40.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 OC Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Bioecker H., Mewes H.W., Lemcke K., Mayer K.F.X., Quetier F.,
 RA Salanoubat M.,
 RU Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RU Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 DR EMBL; AL132979; CAB62435.1; -.
 DR InterPro; IPR001128; -.
 DR Pfam; PF00067; P450; 1.
 DR PRINTS; PR00385; P450.
 DR PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_1.
 KM Heme; Monooxygenase; Oxidoreductase.
 SQ SEQUENCE 513 AA; 58867 MW; B1639BDD9A5D7C93 CRC64;

Query Match 19.6%; Score 503; DB 10; Length 513;
 Best Local Similarity 27.6%; Pred. No. 1,7e-25;
 Matches 144; Conservative 95; Mismatches 216; Indels 66; Gaps 14;

```

OY 4 YTLMTFLCTIVLPVLFLAANKLWEMLMIRVDPNCRSLPFGTGLPFGTGLTQI-- 61
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
DB 7 HTLLPLFLPSLLSLFL-----ILTKRRNRKTRFNLPPKSGMPLFGTIGYLR 58
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
OY 62 ---LQRRKFLMRKQKCYTKTHLEGNPTVRMGADNVQIILGEIKLYSVQMPASVRT 118
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
DB 59 YTTTTLADFEQOQVSKGKTYRSLTFEPITVSADAGLNFILQNEGRLECSYPRSIG 118
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
OY 119 ILGSDTLNVNGVOHKKKKRMRKAFSRDALEH-----YIPVQOEKSAIOEWLQKDS 172
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
DB 119 ILGKMSMLVYVGMHND-----MRSISLNFSLHARLRTILTKDYERTLLVLDQM-QONS 172
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
OY 173 CVLVYEMKLMRIARMLILGFEPEQIKTDEBELVAFEMIKNLSLPIDVPSGLYR 232
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
DB 173 ISSADQAEKFTNLMKHHMSMDPGEELT--POLKKEYTFEMKGVVSADLNLPGTAYHK 230
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
OY 233 GLRAR-----NFIHSKTEENIRKKIODDNEQKYDALQL-----IENSRSDE----- 279
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
DB 231 ALQSRATILKFIKMKKEERKLDIKEDQEEVEKTEDEAMSKSDHYRKQRTDDLLGWY 290
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
OY 280 ---PPLQAMKEATELLFGHETTASTATSLVMFLGLNTEVVOKVRER---VOEKVE 331
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
DB 291 LKHSNLSTEOIDLILSLFAGHETSSAVALAIFLQACPRAVEELREHLEIARAKKE 350
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
OY 332 MGYTPFGKGLSMELDLQKTYGCVIKETLRINPVPYGGFVALKTELNGYQIPKGMV 351
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
DB 351 LG-----ESELNMDQYKMPOTCVINETLRNLGNVFLRHLKALKDVRKYKYDIPSGKYL 406
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
OY 392 YSICDTHVADVPFNKEEPOPERFM-----SKGLEDSRF--NYIPFGSGRMCGK 442
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
DB 407 PYISAVHLNLSRNDQNLNPNRMWQOONNCGASSGSPSTWNMMPTFGGGRRLCAGSE 466
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
OY 443 FAKVLKILVELTQHCNMLISNGPPTMKGTPTIYVDNLP 483
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
DB 467 LKLEMAVFIHHLVLFKNWELAEDEK-----PFAFPFVDFP 502
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :

```

RESULT 14
 065624 PRELIMINARY; PRT; 457 AA.
 AC 065624;
 DT 01-AUG-1998 (Tremblrel. 07, Created)

DT 01-AUG-1998 (Tremblrel. 07, Last sequence update)
 DE 01-MAR-2001 (Tremblrel. 16, Last annotation update)
 DR CYTOCHROME P450
 GN T18B16.200 OR AT4G19230.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 OC Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Bevan M., Benes V., Rechmann S., Borkova D., Ansoerge W., Bancroft I.,
 RA Mewes H.W., Mayer K., Schueller C.;
 RU Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Meyers V., Rechmann S., Borkova D., Ansoerge W., Mewes H.W., Lemcke K.,
 RA Mayer K.F.X.;
 RU Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 131-457 FROM N.A.
 RA Van Der Schueren J., Chuang T.J., Aert R., Defoor E., Robben J.,
 RA Volckaert G., Mewes H.W., Lemcke K., Mayer K.F.X.;
 RU Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RU Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 DR EMBL; AL021687; CA16713.1; -.
 DR HSSP; P33006; CAB78925.1; -.
 DR InterPro; IPR001128; -.
 DR Pfam; PF00067; P450; 1.
 DR PRINTS; PR00385; P450.
 DR PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_1.
 KM Heme; Monooxygenase; Oxidoreductase.
 SQ SEQUENCE 457 AA; 52436 MW; 6105FB7C181EAF07 CRC64;

Query Match 19.1%; Score 489.5; DB 10; Length 457;
 Best Local Similarity 28.2%; Pred. No. 1.1e-24;
 Matches 127; Conservative 97; Mismatches 183; Indels 43; Gaps 12;

```

OY 20 LFLAANKLW--EMLMIRVDPNCRSLPFGTGLPFGTGLTQI-Q-RRKFLMRKQKYG 76
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
DB 10 LFGASLFLYLRCLISRRRGSSKPLRPSTMGPIYGETFOLYSQDPNVFPOSKQRRYG 69
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
OY 77 CIYKTHLEGNPTVRMGADNVQIILGEIKLYSVQMPASVRTILGSDTLNVNGVOHKNR 136
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
DB 70 SVFKTHVLCPCVMISSPEAKFVLVTKSHLFPKTPASKRMIGKQAIFFHGGDYHAKL 129
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
OY 137 KRAIMAFSRDALEHNYIPVQOEKSAIOEWLQKDSVLVYPPMKLMRIARILGFE 196
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
DB 130 KILVLFNPESTRNMPDLESTAODSLKSM--EGVINTYQEMKTYTFVALLSLFG-- 185
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
OY 197 PROIKTDE-----OELVAFEMIKNLSLPIDVPSGLYGLRARNFIHSKTEENIRKKI 252
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
DB 186 ----KDEVLYREDLKRKYTYLLEKGYNSMPVNLPGTLFHKSMAR-----KELSOILARI 235
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
OY 253 QDDNNEQKYKAL-QLIENSRSDEPSPLOAMKEATELLFGHETTASTATSLVMF 311
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
DB 236 LSERRONGSSHNDLGSFMDKKEELTDEQIA-----DNIIGVIFAARDTASVMSWLK 290
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
OY 312 LGNTEVVOKVRER---OEKVEGMGYTPFGKGLSMELDLQKTYGCVIKETLRINPVP 367
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
DB 291 LAMPNVLEAVTEQMAIKRQKEE-----GESLIMGDTKKMLPISRVLQETLRVSIIS 344
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
OY 368 GGFVALKTELNGYQIPKGMVYISICDTHVADVPFNKEEPOPERFMFGSKGLEDSRFN 427
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
DB 345 FTFREAVEDEVVEGYLIPKGMKYLPLFRNLIHSDADIFSNGKRDPSRF-----EVAPKRN 399
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
OY 428 -YIPFGSGRMCGKREFAVLLKILVELT 456
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :

```


Db 400 TEMPEGNGTHSCPGENLAKLEMSIMHHLT 429

Search completed: November 6, 2001, 13:38:25
Job time: 253 sec

RESULT 15

Q9FMA5 PRELIMINARY; PRT; 465 AA.
ID Q9FMA5
AC Q9FMA5;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE CYTOCHROME P450.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N. A.
RC STRAIN=COLUMBIA;
RX MEDLINE=98290546; PubMed=9628582;
RA Sato S., Kaneko T., Kotani H., Nakamura Y., Asamizu E., Miyajima N.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. IV.
RT Sequence features of the regions of 1,456,315 bp covered by nineteen
RT physically assigned P1 and TAC clones."
RL DNA Res. 5:41-54(1998).
DR EMBL: AB009048; BAB08653.1; .
SQ SEQUENCE 465 AA; 53767 MW; 0C00459C9C866D1F CRC64;

Query Match 18.88; Score 481.5; DB 10; Length 465;
Best Local Similarity 26.9%; Pred. No. 3.8e-24;
Matches 130; Conservative 110; Mismatches 197; Indels 47; Gaps 13;

QY 1 MGLYTLMTFTCTVLYPLVFLAVALKLEMLIRVDPNCRSPPTGTPGTGETLQL 60
DB 8 MGLIILIVS-LCSALL-----RNNQRY-----TKNGLPPTGMPPIGETTEF 50
QY 61 ILQRRKFLMKRKKYGYCTHLEFNGPTVRVAGADNRQILLGEHKLVSQWPAVSRTIL 120
DB 51 LKQGNFMFRNQRLRGYGFSGHLLGCPPLISDSEVNNRYLLKNEKGLVPGYQSMIDL 110
QY 121 GSDTLNWHGVQHKKKAIRAFSRDAL-EHYIPVIOEYKSAIQEWLQKDSVLYYPE 179
DB 111 GTCNMAAVHGSSHRMLRGLSLSLISSTMARDHILPKVDHFMKSYLDQWNELE-VIDIQDK 169
QY 180 MKLMFRTAMRLLLGPEPQIKTDEQLEAFEEEMIKNLFSLPIDVPFSGLYRGIRARNF 239
DB 170 TKHMAFLSLTIQAG---NLKRPVEEFKTAFFKLVGTLSPIDLPGTINRGIGIQRANN 226
QY 240 IHSKTEENIRKKIODDNEENQKYDALQLLI-ENSRSDPEFSLQMKKEATELLFEG 297
DB 227 I-----DRLRLMOERDRSGETFTDMLGYLMKKEGNR---YPLFDEIRDQVVTILYSG 278
QY 298 HETASTATSLVMTGLNTEVQKXREE---VQEKVEMGMTPPGKLSMELLDOIKYTG 354
DB 279 YETVSTTSMALKYLDHDKALQELRAELAFERKKRQ-----DEPLGLEDVKSMKFTRA 333
QY 355 VIKETLRINPVPYGGFRVALKTFELNGYOIPKGMVIVISICDTHDVADEVFPNKEEFOPER 414
DB 334 VIEYERLRTIYNGVLRKTRTROLEINGYILPRGWRIVYTRREINDANLYEDPLIFNPMR 393
QY 415 FMSKGLDGSRRNYIPGGGSRMCYKKEPAKYLKILVELTQHCNWLISNGPPTMKTGP 474
DB 394 WMKKSLE--SONSCFVFGGTRICPKELGIVEISSFLHYFVTRYRMWEIIGDELM--- 447
QY 475 TIYP 478
DB 448 -VFP 450

OM of: US-09-668-482-2 to: GenEmbl:* out_format : pfs
Date: Nov 6, 2001 3:52 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODE=frame+pzr.model -DEV=xlp
-O=/cgn2.1/USPT0.spool/US09668482/runat.05112001.133353.10884/app-query.fasta.1.1680
-DB=GenEmbl -OEMT=fastap -SUFFIX=rg -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPT=0.000 -LOOPTXT=0.000 -GAPOP=4.500
-GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -YGAPEXT=10.000 -YGAPEXT=0.500 -DELOP=6.000
-LIST=45 -DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OEMT=pfs -NORM=ext -MINLEN=0
-MAXLEN=200000000 -USER=US09668482.ccgcnl.1.8677 -NCPU=6
-ICPU=3 -LOGLOG -NO_XLPPX -WAIT -THREADS=1

Search information block:

Query: US-09-668-482-2

Query length: 492

Database: GenEmbl:*

Database sequences: 1344157

Database length: -85606004

Search time (sec): 5019.990000

score_list:

Sequence	Strd Orig	zScore	Escore	len	Documentation
gb_ov:DRU68234	+ 2563.00	3135.38	2.1e-166	1479	U68234 Danio rerio all-trans-7
gb_ov:AF1199462	+ 1816.50	2221.26	1.7e-115	1479	AF1199462 Gallus gallus retinol
gb_ov:AF005418	+ 1741.00	2127.23	3.0e-110	1743	AF005418 Homo sapiens retinol
gb_ov:AF115769	+ 1738.00	2123.65	4.7e-110	1726	AF115769 Mus musculus cytochc
gb_ov:AF05509A	+ 1734.00	2118.59	9.0e-110	1756	AF05509A Mus musculus RNA for p45
gb_ov:AF057566	+ 1646.50	2013.08	6.8e-104	1479	AF057566 Xenopus laevis retinol
gb_hvg21:AL358613	+ 1279.00	1517.85	2.6e-76	166337	AL358613 Homo sapiens chrom
gb_ov:AF252297	+ 1034.50	1253.12	1.5e-61	4445	AF252297 Homo sapiens cytochc
gb_ba3:YSC51E	+ 953.50	607.62	1.4e-36	113064	D64003 Synchocystis sp. PC
gb_pat1:AR074363	+ 552.50	672.62	3.1e-29	1608	AR074363 Sequence 1 from patent
gb_pat1:ATCVP450R	+ 528.50	643.93	1.2e-27	1494	AF118211 Taxus cuspidata 5-ald
gb_pat1:AF218211	+ 515.50	627.07	1.1e-26	1649	AF218211 Vigna radiata cytochc
gb_pat1:AF279252	+ 489.00	585.88	1.3e-22	116205	AF279252 Arabidopsis thalian
gb_pat1:AC006931	+ 481.00	568.42	2.0e-24	1395	U54770 Solanum lycopersicum cy
gb_ov:AF1854770	+ 460.00	574.74	8.9e-24	322	AF185477 Gallus gallus CYP26 (C
gb_ov:AB008097	+ 455.00	551.46	1.8e-22	1934	AB008097 Arabidopsis thaliana
gb_pat1:AF318500	+ 435.50	547.85	2.8e-22	1487	AF318500 Arabidopsis thaliana
gb_pat1:AF212991	+ 430.00	528.71	3.2e-21	1718	AF212991 Cucurbita maxima ent
gb_pat1:AF218112	+ 430.00	486.42	7.4e-19	70475	AF218112 Arabidopsis thaliana
gb_pat1:AF318501	+ 394.50	478.38	2.1e-18	1740	AF318501 Arabidopsis thaliana
gb_pat1:AF326277	+ 384.00	466.97	1.2e-17	1844	AF326277 Hordeum vulgare ent-X
gb_pat1:AB020744	+ 377.50	423.35	2.4e-15	62052	AB020744 Arabidopsis thaliana
gb_pat1:ZMU32579	+ 375.00	449.59	4.4e-17	1724	U32579 Zea mays DWARF3 (dwarf3
gb_ba3:KMTAR	+ 370.50	449.33	8.6e-17	1681	AF232955 Myxococcus xanthus cy
gb_pat1:AT118B16	+ 363.50	401.97	3.7e-14	96574	AF2021687 Arabidopsis thaliana
gb_pat1:ATCCHR1V50	+ 363.50	395.04	9.1e-14	199382	AF161550 Arabidopsis thaliana
gb_ba3:KMTCY22D7	+ 363.00	411.98	1.0e-14	31859	Z881866 Mycobacterium tubercu
gb_pat1:AF251548	+ 359.00	436.13	4.7e-16	1533	AF251548 Tribolium castaneum c
gb_ov:AM966077	+ 350.50	424.17	2.2e-15	1802	AF26077 Oncohyacinthus mykiss c
gb_pat1:AF280108	+ 347.00	421.56	3.0e-15	1512	AF280108 Homo sapiens clone 15
gb_pat1:AF319634	+ 347.00	421.56	3.0e-15	1512	AF319634 Homo sapiens cytochc
gb_ov:GGA250337	+ 346.00	419.47	5.1e-15	1638	AF250337 Gallus gallus myrnat fc
gb_pat1:AB017785	+ 346.00	417.57	5.1e-15	2041	AB017785 Mus musculus cyp4A10
gb_ba3:SC7E4	+ 344.50	387.18	2.3e-13	36772	AF159214 Streptomyces coelic
gb_ov:RASC450PA	+ 343.50	415.18	6.8e-15	1882	AF159214 Rabbit CYP3A6 gene enc
gb_ov:AF151272	+ 343.50	413.19	8.8e-15	2316	AF151272 Oryzias latipes cyto
gb_pat1:AF166266	+ 343.00	414.84	7.2e-15	1830	AF166266 Mus musculus lanost
gb_pat1:AF139532	+ 343.00	414.56	7.4e-15	1883	AF139532 Liliandamar styractif
gb_pat1:AC007002	+ 342.00	372.07	1.7e-12	140331	AF195532 Homo sapiens BAC cl
gb_pat1:E02279	+ 340.50	411.71	1.1e-14	1843	E02279 Human liver cytochrome

gb_pat1:AB037244	+ 339.50	412.42	9.8e-15	1506	AB037244 Asparagus officina
gb_pat1:AB037245	+ 339.50	412.42	9.8e-15	1506	AB037245 Asparagus officina
gb_ov:RABP4503C	+ 339.50	410.88	1.2e-14	1768	M19139 Rabbit cytochrome P4
gb_pat1:HM0CYP209A	+ 339.50	410.47	1.3e-14	1845	M61857 Human cytochrome P45

seq_name: gb_ov:DRU68234

seq_documentation block: 1479 bp mRNA VRT 22-NOV-1996
LOCUS DRU68234
DEFINITION Danio rerio all-trans-retinoic acid 4-hydroxylase (p450RAI) mRNA,
complete cds.

ACCESSION U68234
VERSION U68234
KEYWORDS U68234.1 GI:1680715

SOURCE zebrafish.
ORGANISM Danio rerio

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Rasbora; Danio.

REFERENCE 1 (bases 1 to 1479)

White,J.A., Guo,Y.D., Baetz,K., Beckett-Jones,B., Bonasoro,J.,
Hsu,K.E., Dilworth,F.J., Jones,G. and Petkovich,M.

Identification of the retinoic acid-inducible all-trans-retinoic
acid 4-hydroxylase

J. Biol. Chem. 271 (47), 29922-29927 (1996)

REFERENCE 2 (bases 1 to 1479)

White,J.A., Guo,Y., Baetz,K., Beckett-Jones,B., Bonasoro,J., Hsu,K.
E., Dilworth,F.J., Jones,G. and Petkovich,M.

Direct Submission
Submitted (28-AUG-1996) Cancer Research Labs, Queen's University,
Bottrell Hall, Kingston, ON K7L 3N6, Canada

JOURNAL
TITLE Direct Submission
AUTHORS White,J.A., Guo,Y., Baetz,K., Beckett-Jones,B., Bonasoro,J., Hsu,K.
E., Dilworth,F.J., Jones,G. and Petkovich,M.

FEATURES
Source Location/Qualifiers

1..1479
/organism="Danio rerio"
/db_xref="taxon:7955"

1..1479
/gene="p450RAI"

1..1479
/gene="p450RAI"

/note="all-trans-retinoic acid 4-hydroxylase"

/protein_id="AAC60045.1"

/db_xref="GI:1680716"

/translation="MGLTYLTVFLYVLLPVLFLAATKMLMIRRVDCRSP
PECTMGLPRIGELIILORRKLPRKROKCYCTHFLGPTRVKMGADNVOILL
GEKLYSVQMPASVRIIGSDPLSNVGHONKKAIRARSDALEYIPVIOEY
KSAIQMDLQDSCVLYVPEKKLMRIARIRILLGDDNENQKTKDALQLITNSR
SDPFSIQMKAEATLRLGHEHTATSTSLVMEFLGTVVAVVREVSQVEMGM
YTPGKLSMELDLKVTGCVIKETLRINPVPVGFRLVAKTFELNGVQIPKGMVY
SIDTTHVDVADVEFERMSKGLSDGSFNTIPGGSRLMCKVGEFAKVLKI
FLVELHONCWILNSGPIPMKGPITYPDNLPTKTSYVRN"

BASE COUNT 425 a 331 c 379 g 344 t

ORIGIN

alignment_scores:
Quality: 2563.00 Length: 492
Ratio: 5.209 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-668-482-2 x DRU68234 ..

Align seg 1/1 to: DRU68234 from: 1 to: 1479

1 MetGlyLeuTyTrhLeuMetValThrPheLeuGlyThrIleValLeuPr 17
|||||
1 ATGGGCTGTACACCTTATGTCACCTTCTCTCACCATCTGCTCTAC 50
|||||
17 ATGAGGCTGTACACCTTATGTCACCTTCTCTCACCATCTGCTCTAC 34
|||||

51 CGTTTACTCTTTCGCCGGGTGAAGTTGTGGAGATGTTAATGATCC 100
 34 rGArValAspProAsnCysArgSerProLeuProGlyThrMetGly 50
 101 GACGAGTGTGATCCGAACTGCAGAACTCCTTACCCGCGAGTACCAATGGGC 150
 51 LeuProPheIleGlyGluThrLeuGluLeuIleLeuGlnArgArgLysPhe 67
 151 TTGCGGTTCATTTGAGAAACGCTCAGCTGATTCCTCCAGAGAGAACTT 200
 67 eLeuArgMetLysArgGlnLysTyrGlyCysIleTyrLysThrHisLeuP 84
 201 TCTGCGATGTAAGCGCAAAATACGGGTGCATCTACAAAGCAGCACTCT 250
 84 heGlyAsnProThrValArgValMetGlyAlaAspAsnValArgGlnIle 100
 251 TCGGGAACCCGACTGTCAAGGTGATGGAGCTGATTAATGTACGACAGATT 300
 101 LeuLeuGlyGluHisLysLeuValSerValGlnTrpProAlaSerValArg 117
 301 CTGCTGGCGCAACACAAAGCTGTCTGTCTCAGTGGCCAGCATCAGTGAG 350
 117 gThrIleLeuGlySerAspThrLeuSerAsnValHisGlyValGlnHisL 134
 351 AACCATCTGGGCTGTGACACCTCTCCAAATGTCCATGAGATGTAACACA 400
 134 ysaAsnLysLysAlaIleMetArgAlaPheSerArgAspAlaLeuGlu 150
 401 AAAACAGAAAAAGCCATTATGAGGGGCTTCTCGAGATGCTCTGGAG 450
 151 HisTyrIleProValIleGlnGlnGluValLysSerAlaIleGlnGluTr 167
 451 CACACATTCCTCCGATGCAGACAGAGGTGAAGCCCATACAGGAAATG 500
 167 pLeuGlnLysAspSerCysValLeuValTyrProGluMetLysLysLeu 184
 501 GCGCAAAAAGACTCCTCGCTGCTGCTTTCAGAAATGAGAAATCA 550
 184 eThPheArgIleAlaMetArgIleLeuLeuGlyPheGluProGluGlnIle 200
 551 TGTTCGGATAGCTATGAGAAATCCTGCTGTTTGAACACAGAACATA 600
 201 LysThrAspGluGlnGluLeuValGluAlaPheGluGluMetIleLysAs 217
 601 AAACGAGCAGCAAGAACTGTGGAAGCTTTGAGGAAATGATCAAAAA 650
 217 nLeuPheSerLeuProIleAspValIleProPheSerGlyLeuTyrArgL 234
 651 CTGTCTCTCTTGCATTCGACCTTCTTCAATGAGTGTCTTACAGGGCTT 700
 234 euArgAlaArgAsnPheIleHisSerLysIleGluGluAsnIleArgLys 250
 701 TGAGGGCAGCAATTTCACTCACTCCAAAATGAGAAAACATCAGAGAG 750
 251 LysIleGluAspAspAsnGluAsnGluGlnLysTyrLysAspAlaLe 267
 751 AAAATTCAAGATGACGACATGAAAGCAAGCAAAATACAAAACGCGCT 800
 267 uGlnLeuLeuIleGluAsnSerArgArgSerAspGluProPheSerLeu 284
 801 TCAGCTGTGATCGAGACAGCAAGAGTAGACACCTTTAGTTAGTTGC 850
 284 IAlaMetLysGluAlaAlaThrGluLeuLeuPheGlyGlyHisGluThr 300
 851 AGCGGATGAAAGAGACGATACAGACCTCTATTGAGAGTCAATGAAC 900
 301 ThrAlaSerThrAlaThrSerLeuValMetPheLeuGlyLeuAsnArg 317
 901 ACGGCGAGACTGCAACTCACTGTGATGTTCTGGGCTGGAACACAGA 950
 317 uValValGlnLysValArgLugluValGlnGluLysValGluMetGly 334
 951 AGTGTGTCAGAAAGTTCAGAGAGAGGTTGAGGAAGTTGAATGGCA 1000

334 eTyrThrProGlyLysGlyLeuSerMetGluLeuLeuAspGlnLeuLys 350
 1001 TGTATACACTGGAAAGGCTTGAATGAGACTGTGGACCACTGAG 1050
 351 TyrThrGlyCysValIleLysGluThrLeuArgIleAsnProProValPr 367
 1051 TACACTGATGTGTATTAAGAGACTCTTGAATTAACACCTCTCTTCC 1100
 367 oGlyGlyPheArgValAlaLeuLysThrPheGluLeuAsnGlyTyrGlnI 384
 1101 CGAGAGATTACAGACGCGACATCAAAACCTTGAATTAATGATGTTACAAA 1150
 384 leProLysGlyTyrAsnValIleTyrSerIleCysAspThrHisAspVal 400
 1151 TTCTTAAGATGGAACGTCATTTACACATCTGTGACACGACCATGTG 1200
 401 AlaAspValPheProAsnLysGluGluPheGlnProGluArgPheMetSe 417
 1201 GCCGACGTCTTCCAAACAAAGAGAGTTCCAGCCGGAGAGATTCATGAG 1250
 417 rLysGlyLeuGluAspGlySerArgPheAsnTyrIleProPheGlyGly 434
 1251 CAAAGTCTGGAGGACGGGTCCAGGTTTAACATCACTCCCTTCGAGAGAG 1300
 434 LysSerArgMetCysValGlyLysGluPheAlaLysValLeuLeuLysIle 450
 1301 GATCAGAGATGTGTGTGGCAAAAGATTGCCAAAGATTACTCAAGATTC 1350
 451 PheLeuValGluLeuThrGlnHisCysAsnTrpIleLeuSerAsnGlyPr 467
 1351 TTTTATGTTGATTAAACGACATTCGAATGATTCCTCAAAAGGACC 1400
 467 oProThrMetLysThrGlyProThrIleTyrProValAspAsnLeuProt 484
 1401 CCGACATGAAACAGCCCGCACTATTACCCAGTGTGACATCTCCCTA 1450
 484 hLysPheThrSerTyrValArgAsn 492
 1451 CCAATTCACACTAGTTATGTCAGAAT 1476

seq_name: gb.ov:AF199462
 seq_documentation_block: 1479 bp mRNA VRT 21-NOV-1999
 LOCUS AF199462 1479 bp mRNA VRT 21-NOV-1999
 DEFINITION Gallus gallus retinoic acid degrading enzyme CYP26 mRNA, complete cds.
 ACCESSION AF199462
 VERSION AF199462.1 GI:6456740
 KEYWORDS
 SOURCE
 ORGANISM
 Chicken.
 Gallus gallus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 Phasianinae; Gallus.
 1 (bases 1 to 1479)
 Swindell,E.C., Thaller,C., Sockanathan,S., Petkovich,M.,
 Jessell,T.M. and Eichele,G.
 Complementary Domains of Retinoic Acid Production and Degradation
 in the Early Chick Embryo
 Dev. Biol. (2000) In press
 2 (bases 1 to 1479)
 Swindell,E.C., Thaller,C., Sockanathan,S., Petkovich,M.,
 Jessell,T.M. and Eichele,G.
 Direct Submission
 Submitted (28-OCT-1999) Max Planck Institute for Experimental
 Endocrinology, Feodor-Lynen Strasse 7, Hannover 30625, Germany
 TITLE
 JOURNAL
 FEATURES
 source
 CDS
 1..1479
 /organism="Gallus gallus"
 /db_xref="taxon:9031"
 1..1479
 /note="P450RAI"


```

seq_name: gb_pr4:AF005418
seq_documentation_block:
LOCUS       AF005418          1743 bp    mRNA                    PRI
DEFINITION  Homo sapiens retinoic acid hydroxylase mRNA, complete cds.
ACCESSION   AF005418
VERSION     AF005418.1   GI:2688845
KEYWORDS
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 1743)
AUTHORS     White,J.A., Beckett-Jones,B., Guo,Y.D., Dilworth,F.J., Bonasoro,J.,
            Jones,G. and Petkovich,M.
TITLE       cDNA cloning of human retinoic acid-metabolizing enzyme (hP450RA1)
            identifies a novel family of cytochromes P450
JOURNAL     J. Biol. Chem. 272 (30), 18538-18541 (1997)
MEDLINE     97373542
REFERENCE   2 (bases 1 to 1743)
AUTHORS     White,J.A., Beckett-Jones,B., Guo,Y., Dilworth,F.J., Bonasoro,J.,
            Jones,G. and Petkovich,M.
TITLE       Direct Submission
JOURNAL     Submitted (26-MAY-1997) Cancer Research Labs, Queen's University,
            Botterell Hall, Rm 355, Kingston, Ont K7L 3N6, Canada
FEATURES
    source   1..1743
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /note="hP450RA1; cytochrome P450"
             /codon_start=1
             /product="retinoic acid hydroxylase"
             /protein_id="AAB88881.1"
             /db_xref="GI:2688845"
             /translation="MGLPRLLSALCTFPLPLLLFLAIAIKLMDYCVSRDSCALPL
             PGTMGFPTGGLTQVILQNRKFLQMKRRKYGFTYKTHLFGAPYVAGADVNRILL
             GDRLVSVHWPASVRITLIGSGLSNLHDSHQKRVIKRAASREALIEGVVITEEV
             GSLDQWLSCGERGLVYEVKRLMFIARILLCEPGLAGDSEQQLVAFEEMT
             RMFLFPLIDVPSGLYRGKARNLIHARLEONIRAKICGLRASEGGQCKDLQILLIE
             HSEMERGLMDQALQSSTELFJGGHTTASATSLITVIGLYPHYLQVREELKSKG
             ILCKSNQNKIDMELIEOLKYICVYIKETLRINPVPVPGSFARALATFEELNGQIPIRGW
             NVIYICDTHVAELFTNKEENNPREFMLPHEDSRESFIFGGGLRSCVKEPAKI
             LKITTVELARHCMDQNLNGPPTMTSPVYVDMLPARTFHHGI"
BASE COUNT  423 a      442 c      477 g      401 t
ORIGIN
alignment_scores:
    Quality: 1741.00      Length: 497
    Ratio: 4.087          Gaps: 5
    Percent Similarity: 85.714      Percent Identity: 68.209
alignment_block:
US-09-668-482-2 x AF005418 ..
Align seg 1/1 to: AF005418 from: 1 to: 1743
1 MetGlyLeuThrLeuMetValThrPheLeuGlyThrIleValLeuPr 17
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
39 ATGGGCTCCCGCGCTGCTGCGCAGTGGCTGTGCACTTCGTGCGTGC 88
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
17 oValLeuLeuPheLeuAlaValIleLeuThrPGLuMetLeuMetIle 34
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
89 GCTGCTGCTCTTCTGCTGCTGCGATCAAGCTCTGGACCTACTGCGTGA 138
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
34 rGArYalAspProAsnCySArGSerProLeuProProGlyThrMetGly 50
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
139 GCGGCGCGACCGCAGTGTGCTCCATTCGCCCGCGGAGACTATGAGGC 188
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
51 LeuProPheIleGlyIuThrLeuGlnLeuIleLeuGlnArGArGlySph 67
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

189 TTCCCTCTTCTGGGAAACCTTGACATGCTGACGCGGAGAAATT 238
67 eLeuArMetLysArGlnLysTyrGlyCysIleTyrLysThrHisLeuP 84
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
239 CCTGCAGATGAAGCGCAGAAATATACGGCTTCACTACAAAGACATCTGT 288
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
84 heGlyAsnProThrValArValIleMetGlyAlaAspAsnValArgGlnIle 100
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
289 TCGGGCGGCCACCGCTACGAGTGATGGCGGACAAATGTGGCGGCATC 338
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
101 LeuLeuGlyGluHisLysLeuValSerValGlnTrpProAlaSerValAr 117
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
339 TTGCTCGGAGACGACCGCGCTGTGCTGCTGCTGCTGCTGCTGCTGCTG 388
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
117 gThrIleLeuGlySerAspThrLeuSerAsnValHisGlyValGlnHisL 134
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
389 CACCATTTCTGGGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 438
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
134 ySAAsnLysLysLysAlaIleMetArgAlaPheSerArgAspAlaLeuL 150
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
439 AGCAGCGCAAGAGGTATATAGCGGCTTCAGCGCGAGGACTCGAA 488
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
151 HisTyrIleProValIleGlnGlnValLysSerAlaIleGlnGlnL 167
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
489 TGCCTACGTGCGCGTATCACCGAGAAAGTGGGACAGCCCTGAGACATG 538
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
167 pLeuGlnLysAspSerCys.....ValLeuValTyrProGlnL 180
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
539 GCTG.....AGCTGCGGCGAGCGCGGCTCTGCTTACCCCGAG 579
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
180 eTylLysLeuMetPheArgIleAlaMetArgIleLeuLeuGlyPheGlu 196
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
580 TGAAGCGCTCATGTTCGGAATCCGCAATCCGCAATCCGCTGCGCA 629
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
197 ProGlnLysIleLysThrAsp.....GluGlnLeuValGluAl 210
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
630 CCC...CAACTGCGGCGGCGGAGCTCCGAGCAGCAGCTTGAGAGGC 676
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
210 aPheGluGlnMetLysLysAsnLeuPheSerLeuProIleAspValProp 227
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
677 CTTCGAGGAATGACCCGCAATCTTCTGCTGCTGCTGCTGCTGCTGCT 726
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
227 heSerGlyLeuTyrArgLysLeuArgAlaArgAsnPheIleHisSerLys 243
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
727 TCACCGGCTGTACCGGCGGATGAAGGCGCGGACCTCATTCACGCGGC 776
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
244 IleGlnLysAsnIleArgLysLysIleGlnAsp...AspAsnGlnL 259
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
777 ATCGAGCAGACATTCGCGCCAGATGCGGCGCTGCGGCGATCCGAGGC 826
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
259 nGluGlnLysTyrLysAspAlaLeuGlnLeuIleGluAsnSerArg 276
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
827 GGGCAGGCGCTGCNAAGACCGCTGCTGCTGCTGCTGCTGCTGCTG 876
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
276 rGSerAspGluProPheSerLeuGlnAlaMetLysGlnAlaAlaThrGlu 292
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
877 AGAGGGAGAGAGCGGCTGACATGACGACACTAAAGCAATCTCAACGAA 926
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
293 LeuLeuPheGlyHisGlnThrThrAlaSerThrAlaThrSerLeuVal 309
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
927 CTCTCTTTGGAGACACGAACACGCGCAGTGCACCACTCTGAT 976
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
309 lMetPheLeuGlyLeuAsnThrGluValValGlnLysValArgGlnL 326
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
977 CACTTACCTGGGCTGACCAACATGTTCTCAAGAAAGTCCGAGAAAGC 1026
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
326 aLglnLysValLysMetGlyMetTyrThrProGlyLysGlyLysSer 342
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1027 TGAAGAGTAAGGTTACTTGAAGAGCAATCAAGACAAAGTTGAGAC 1076
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
343 MetGluLeuLeuAspGlnLeuLysTyrThrGlyCysValIleLysGlu 359
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1077 ATGGAATTTTGAACAACTTAATACATCGGCTGTGTATTATAGGAGAC 1126
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

359 rleuarglleasnproprovalprogllygylpheagvalalaleuylst 376
      |||
1127 CMTTCGACGTAAATCCCAAGTTCAGAGGGTTTCGGGTGCTCGAAGA 1176
      |||
376 hrhhegluleuasnnglytrglnlleprolysglytrpaspvalilletyr 392
      |||
1177 CTTTGTGATTAATAGGATACAGATTCACCAAGGCTGGAAATGTTATCTAC 1226
      |||
393 SerlleCysaspThrhIsaspValAlaaspValpheproasnlysglucl 409
      |||
1227 AGATATCGTATACATGATGTGGAGAGATCTTCACCAACAAGAGA 1276
      |||
409 upheglinprogluarphemetserlysglyleuGluaspGlySerArp 426
      |||
1277 ATTTAAATCCGACCGATTCATGCTGCTCACCACCAAGAGATGATCCAGGT 1326
      |||
426 heasnyrilleprophelglyglySerarpmetyValglylysglu 442
      |||
1327 TCAGCTTCATTCATTTGGAGAGAGCCTTAGAGCTGTGTAGGCAAAAGA 1376
      |||
443 phealalysvalleuLeuLysIlephleuValGluLeuThrGlnHisCy 459
      |||
1377 TTTGCAAAAATTTCTTCAAAATATTATACAGTGAGCTGGCCAGGCAATG 1426
      |||
459 sasnrTrpIleuSerasnnglyproprothrmetyThrGlyProthri 476
      |||
1427 TGAGTCGACGCTCTTAATAGACCTCCTACCAATGAAACCAAGTCCCAAG 1476
      |||
476 letyrProvalaspasnleuprothrmetyThrLysPheThrSerTy 489
      |||
1477 TGTATTCGTGTGACATCTCCCTCCAGATTCACCATTTTC 1517
      |||
seq.name: gb_f01:AF115769

```

```

seq_documentation_block:
LOCUS      AF115769      1726 bp      mRNA      03-MAR-1999
DEFINITION Mus musculus cytochrome P450 retinoic acid metabolizing enzyme
ACCESSION AF115769
VERSION    AF115769.1 GI:4324987
KEYWORDS   .
SOURCE     house mouse.
ORGANISM   Mus musculus.
            Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 1726)
AUTHORS   Paine,C.T., Paine,M.L. and Snead,M.L.
TITLE     Identification of amelogenin- and tufel1in-interacting proteins
          using the yeast two-hybrid system
JOURNAL    Connect. Tissue Res. 38, 257-267 (1998)
REFERENCE  2 (bases 1 to 1726)
AUTHORS   Paine,C.T., Paine,M.L. and Snead,M.L.
TITLE     Direct Submission
JOURNAL    Submitted (21-DEC-1998) CCMB, University of Southern California,
          2250 Alcazar Street, CSA103, Los Angeles, CA 90033, USA
FEATURES   location/Qualifiers
            1..1726
            /organism="Mus musculus"
            /db_xref="taxon:10090"
            /tissue_type="tooth"
            /dev_stage="1 day old"
            27..1520
            /note="tufel1in-interacting protein 24 T1P-24"
            /product="cytochrome P450 retinoic acid metabolizing
            enzyme P450RA"
            /protein_id="AA017217.1"
            /db_xref="GI:4324988"
            /translation="MGIPLALATLCTFVPLILFLAALIKIMLYCVRSDRCALPL
            PGTMGPFPGETLQVLRKRFLOMRKRTIGTYKTHLEGRPTVYRMGDNVRRILL
            GHRLVSHVMASVRIILGAGCSLNLSHKKPRKRYKIMQFSGREALQCVLVAEEV
            SSCLEQWLSGGERGLIVPEVKRIMFIARILLGCRPGAGSGEDEQQLVEAFEEET

```

```

BASE COUNT      405 a      428 c      476 g      417 t
ORIGIN
alignment_scores:
      Quality: 1738.00      Length: 496
      Ratio: 4.080      Gaps: 4
      Percent Similarity: 85.887      Percent Identity: 67.944
alignment_block:
US-09-668-482-2 x AF115769 ..
Align seg 1/1 to: AF115769 from: 1 to: 1726
1 MetGlyLeuTyThrLeuMetValThrpheLeuCyThrIleValLeuPr 17
      |||
27 ATGGGGCTCCCGCGGCTGCTGGCACTGCGCTGACCTTGACCTTGCTGCC 76
      |||
17 ovalleuLeuPheLeuAlaValLysLeuTrpGluMetLeuMetIleA 34
      |||
77 GCTGCTGCTCTTCTGCGCGCGCTCAAGCTTGAGACTGTACTGTGTA 126
      |||
34 rgarValaspProasnCysArgSerProLeuproProGlyThrmety 50
      |||
127 GCAGCGCGCATCGCACCTGCGCCCTCCCTTGCCCGCGGTACACAGGC 176
      |||
51 LeuProPheIleGlyIuThrLeuGlnLeuIleLeuGlnArgArgLysPh 67
      |||
177 TTCCCATTTCTTTGGGAAACATTCAGATGCTTCACGCGAGAACTT 226
      |||
67 eleuArgMetLysArgGlnLysTyrgLysIleTyrlLysThrHisLeuP 84
      |||
227 TCTGCAGATGAAGCGCAAAATAGCGCTTCATCTACAAAGCACACTGT 276
      |||
84 heglYasnProthrValArgValMetGlyAlaaspasnValArgGlnIle 100
      |||
277 TTGGGCGCGCCACGCGTCCGGGTATGAGCGCGGATATGTCGGCCATC 326
      |||
101 LeuLeuGlyGlnHisLysLeuValSerValGlnTrpProAlaSerValAr 117
      |||
327 TTGCTGGAGAGACACCGGTGTGTGCTGCTGACGCGCCGCTGGTGG 376
      |||
117 gThrIleLeuGlySeraspThrLeuSerasnValHisGlyValGlnHis 134
      |||
377 CACCATCTGGGCGCTGCTGCTCTCCAACTGACAGATTCTTCGCACA 426
      |||
134 ysasnLysLysAlaIleMetArgAlaPheSerArgaspAlaLeuGlu 150
      |||
427 AGCAGCGAAGAAAGGTATATAGAGCCTTCAGCGCGAGGACATCCAG 476
      |||
151 HisTyrlleProvalIleGlnGlnValLysSerAlaIleGlnGlnTr 167
      |||
477 TGTACTAGTGTCTCGATCGCTGAGGAAGTCAGCATGTGTGGAGCAGT 526
      |||
167 pheGlnLysaspserCys.....ValLeuValTyTrpGluM 180
      |||
527 GCTA.....AGCTGCGGCGAGCGCGGCTCTCTGCTACCCGAGG 567
      |||
180 etLysLysLeuMetPheArgIleAlaMetArgIleLeuLeuGlyPheGlu 196
      |||
568 TGAAGCGCTCATGTTCGCGCATGCGCATGCGCATGCTGGGCTGGAG 617
      |||
197 ProGluGlnIle.....LysThrAspGluGlnLeuValGluAlaPh 211
      |||
618 CCGGCTCCAGCGCGCGCGGAGAGAGAGCAGCAGCATCTGGAGGCTTT 667
      |||
211 eGluGlnMetIleLysasnLeupheserLeuProIleAspValProPhes 228
      |||

```

```

668 CGAGGAGATGACCGCAATCTCTCTCTCCATGACCTGCCCTTTA 717
228 erglyleutyrrargglyleuargalaargaspheilehisserlysile 244
|||||.....
718 GCGGCCGTGTACCGGGCGGTGAAGGCGGGAACCTTATACAGCGCGCATC 767
245 GluculunsniIeArglylsIleGlnasp... AspaspangluasnI 260
|||||.....
768 GAGAGAGACATTCGCCGCCAGATCCGCCGCTTACAGAGCCGGA 817
260 uGlnlystYrlysaPaIaleuGlnleuIleGluasnSerArgArgS 277
|||||.....
818 TGGGGGTTTGAAGGACCGCTGCAGCTCTCATGAGACACTCGTGGAGA 867
277 eraspjuprProPheSerleuGlnalameIylsgIuaIaIaIaThGluLeu 293
|||||.....
868 GGGGAGAGAGCGCTEGATATCGAGGCACTAAACATATGTCACAGACCTC 917
294 leupheGlyIynIsgIutnThraIaSerThraIaThSerleuValme 310
|||||.....
918 CTCCTTGGTGTATGAAGTACAGCCAGTGTCTCCGACATCAGATGAC 967
310 tPheleuGlyLeuasnThrcIuValaIaGlnlysaValaArgGluValG 327
|||||.....
968 TTACCTAAGACTCTACCCACATCTCTCCAGAAAGTTGAGAAAGATTA 1017
327 InaIuLySaValGluMetYrThrProGlyLysGlyLysSerMet 343
|||||.....
1018 AGAGCAAGGCTTACTTTCAGACAGCAATCAAGACAAAGTTAGACATG 1067
344 GluLeuLeuaspGlnLeuLysYrThrGlyCysValIleLysglutnhrle 360
|||||.....
1068 GAAACTTTGGACAGCTTAATATCATTTGGGTGTCTAAGAGACCT 1117
360 uArgIleasProProValaProGlyIyPheArgValaIaLeuLysThrP 377
|||||.....
1118 GCGATGAACTCCCGCTCCAGAGAGGTTTGGGTGCTCTAAGACCTT 1167
377 heGluLeuasnGlyYrGlnIleProLysGlyTrpaspValIleYrSer 393
|||||.....
1168 TTGAGCTGATGATGATACAGATCCCAAGGCTGGAATGTTATTACGT 1217
394 IieCysAspThrHisaspValaIaaspValaPheProAsnLysglutnhr 410
|||||.....
1218 ATCTGTGACACCCAGATGTGGCAGATATCTTCACTAACAGAGGAAAT 1267
410 eGlnProGluArgPheMetSerLysGlyLeuGluaspGlySerArgPhea 427
|||||.....
1268 TATCCCGACCGCTTATAGTGCCTCATCCAGAGATGCTCCGGTTCA 1317
427 snTyrlIeProPheGlyGlySerArgMetCysValaGlyLysgluPhe 443
|||||.....
1318 GCTTCTTCTCATTTGGAGAGGCTTCCGAGCTGTAGGCAAAAGATT 1367
444 AlaIysValleuLeuysIlePheLeuValGluLeuThrGlnHisCysAs 460
|||||.....
1368 GAAAAAATTCCTTAAAGATATTACAGTGAAGCTGGCTAGGACAGTGA 1417
460 nTrpIleLeuSerasnGlyProProThrMetLysThrGlyProThrIle 477
|||||.....
1418 TTGGCAGCTTCTTAATGACCTCTACATGAAGAACAGCCCACTGTGT 1467
477 yTrProValaspasnleuprOthrlYsPheThSerYr 489
|||||.....
1468 ACCCTGTGACATCTCCCTGCAAGATTCACCTACTTC 1505
seq_name: gb_f01:MMP450RA
LOCUS MMP450RA 1756 bp mRNA R08 08-JAN-1998
DEFINITION M.musculus mRNA for p450RA protein.
ACCESSION Y12657
VERSION Y12657.1 GI:2765213

```

```

KEYWORDS p450RA gene.
SOURCE house mouse.
ORGANISM Mus musculus.
REFERENCE 1 (bases 1 to 1756)
AUTHORS Fujii,H., Sato,T., Kaneko,S., Gotoh,O., Fujii-Kuriyama,Y.,
Osawa,K., Kato,S. and Hamada,H.
Metabolic inactivation of retinoic acid by a novel p450
differentially expressed in developing mouse embryos
JOURNAL Unpublished
AUTHORS Hamada,H.
TITLE Direct Submission
AUTHORS Submitted (17-APR-1997) H. Hamada, Osaka University, Institute for
JOURNAL Mol. & Cell. Biol., 1-3 Yamada-oka, Suita, Osaka, 565, JAPAN
FEATURES
source 1..1756
/organism="Mus musculus"
/strain="C3H"
/db_xref="taxon:10090"
/cell_line="P19"
57..1550
/gene="p450RA"
57..1550
/gene="p450RA"
/function="metabolises retinoic acid"
/codon_start=1
/product="p450RA protein"
/protein_id="CAA73206.1"
/db_xref="GI:2765214"
/db_xref="SWISS-PROT:O55127"
/translation="MDPALASALCTVLPVLEFLAALIKLMDLYCVSSRDSRSCALPL
PPTNGVFPFEEITLQVLRKFLQMKRKGFIYKTLFGPVRVAGADNVRIL
GEHRLVSHPWASVPTIIGAGLSIMHDSHOKKYLMOAFSEALQCYLVIAEY
SSCLEOMVSCERGLLYPEVKRLMFLAMRLIGCEPGAGGDEGOLVAFEMT
RNLESLPIDVPSGLRGVKAENLHARIEENIRAKIRLQATEPDGCKNALOLITE
HSMERGEKIDMOALKOSSTELLFGHETTAASLITFYGLIYPRYLQKVEEISKG
LICKSNDONKIDMETLEBOLKTYIGYIKETLRLNRPVPGFVALKTEFLNQIPLKM
NYTISCTHVDADITFKKEENPDRETVPHPEDSRSFTLPPGGLSRCSVKEPAKI
LKLIPTVELARHCWQMLNGPPTMTSPVYPVNLPRAFYFQGD1"
BASE COUNT 408 a 434 c 494 g 420 t
ORIGIN
alignment_scores:
Quality: 1734.00 Length: 496
Ratio: 4.070 Gaps: 4
Percent Similarity: 85.887 Percent Identity: 67.742
alignment_block:
US-09-668-482-2 x MMP450RA
Align seg 1/1 to: MMP450RA from: 1 to: 1756
1 MetGlyLeuTyThrLeuMetValIhrPheLeuCysThrIleValLeuPr 17
|||||.....
57 ATGGGCGCTCCCGGCGCTGCTGCGCACTGCGCTGCACCTGTGCTGCGC 106
17 oValleuLeuPheLeuAlaValIysLeuTrpGluMetLeuMetIleA 34
|||||.....
107 GCGGCTGCTTCTCGCGCGCTCAAGCTCTGGACCTGACTGTGTGA 156
34 rGArgValaspProAsnCysArgSerProLeuProProGlyThrMetGly 50
|||||.....
157 GCAGCGCGCATGCGACAGCTGCCCTCCCTTGGCCCCCGCGTACATGGGC 206
51 LeuProPheIleGlyGluThrLeuGlnleuIleLeuGlnArgArgIysBh 67
|||||.....
207 TTCCCATCTTTGGGGAAACATTTGAGATGTCTTCAACGAGGAAGATT 256
67 eLeuArgMetLysArgGlnLysTyThrGlyCysIleTyThrHisLeuP 84
|||||.....

```



```

257 TCTGCAGATGAGGCGCAGAAATACGGCTTCATCTACAAAGACGCAATCTG 306
84 heglYasPnroThrValArgValMetGluAlaAspAsnValArgGlnIle 100
TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT
307 TTGGGGCGCCACGCGTCCGGGTGATGGCGCGGATATGTCCGGCGCATC 356
101 LeuGluGlyGluHisLysLeuValSerValGlnTrpProAlaSerValAr 117
TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT
357 TTGCTGGAGAGACCGGTGTGTGTCGTCACATGCGCCCGCTGGGTGCG 406
117 gThIleGluGlySerAspThrLeuSerAsnValHisGlyValGlnHisL 134
TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT
407 CACCATCTGGCGCTGGCTCCCTCCAACTCCACGATTCCTCCGACA 456
134 ySAsnLysLysLysAlaIleMetArgAlaPheSerArgAspAlaLeuGlu 150
TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT
457 AGCAGCCAAAGAGGTGATATGACGCGCTTCACGCCGAGGCGCATCAG 506
151 HisTrpIleProValIleGlnGlnValLysSerAlaIleGlnGluTr 167
TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT
507 TGTACGCGCTCTGATCGCTGAGAGATCAGAGTTGCTGGAGCAGTG 556
167 PleuGlnLysSperCys.....ValLeuValTrpProGlu 180
TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT
557 GCTA.....AGCTGCGGCGAGCGCGGCTCTGCTACCCCGAGG 597
180 eLysLysLeuMetPheArgIleAlaMetArgIleLeuLeuGlyPheGlu 196
TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT
598 TGAAGCGCTCATGTCCGATCCGCATGCGCATGCTGCGTGGCGAG 647
197 ProGluGlnIle.....LysThrAspGluGlnGluValGluAlaPh 211
TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT
648 CCGGGTCCAGCGCGCGCGGAGAGAGCAGACAGCTCGTGAGGCTT 697
211 eGluGluMetIleLysAsnLeuPheSerLeuProIleAspValProPhe 228
TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT
698 CGAGAGAGATGACCCGCATCTCTCTCTCCATTCACATGACGCTGCTTA 747
228 eArgLysLeuTrpArgGlyLeuArgAlaArgAsnPheIleHisSerLysIle 244
TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT
748 GCGGCTCTACCGGCGCGTGAAGCGCGCAACTTATACGCGCGCATC 797
245 GlnGluAsnIleArgLysLysIleGlnAsp...AspAsnGlnLysGlu 260
TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT
798 GAGGAGAACATTCGCGCAGATCCGCGCTTCAGGCTACAGACCGGGA 847
848 TGGGGGTGGCAAGGAGCGGCTGCACCTCTGATGACACCTCGTGGAGA 897
260 uGlnLysTrpLysAspAlaLeuGlnLeuIleGluAsnSerArgArgS 277
TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT
277 eArgPgluProPheSerLeuGlnAlaMetLysGluAlaAlaThrGluLeu 293
TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT
898 GCGGAGAGAGGCTGATATGACAGGCACTAAACAATCGTCAACAGAGCTC 947
294 LeuPheGlyGlyHisGluThrThrAlaSerThrAlaThrSerLeuValMe 310
TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT
948 GTCTTTGGTGATGAACCTACACCGCATGCTCGGACATCCTATCATCAG 997
310 tPheLeuGlyLeuAsnThrGluValValGlnLysValArgGlnGluValG 327
TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT
998 TTACATGAGACTCTACCCACATGCTCCAGAAAGTGCAGAAAGAGATTA 1047
327 InGlnLysValGluMetGlyMetTrpThrProGlyLysLysLeuSerMet 343
TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT
1048 AGGAGCAAGGCTTACTTTCAGAGCAATCAAGACAAAGTAGTAGCATG 1097
344 GlnLeuLeuAspGlnLeuLysTrpThrGlyCysValIleLysGluThrLe 360
TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT
1098 GAAACTTTGGAACAGCTTAATACATGGGTGTGATTAAGAGACCTT 1147
360 uArgLLeasnProProValProGlyLysPheArgValAlaLeuLysThrP 377
TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT
1148 GCGATTGAATCTCTCGGTCGAGAGGTTTGGGTTGCTCTGAAGACTT 1197

```

seq_name: gb_ov:AF057566

seq_documentation_block:

LOCUS AF057566 1479 bp mRNA VRT 10-MAR-1999

DEFINITION Xenopus laevis retinoic acid converting enzyme mRNA, complete cds.

ACCESSION AF057566

VERSION AF057566.1 GI:3283171

KEYWORDS

SOURCE

ORGANISM

African clawed frog.

Xenopus laevis

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;

Xenopodinae; Xenopus.

1 (bases 1 to 1479)

Holleman, T., Chen, Y., Grunz, H., and Pieler, T.

Regionalized metabolic activity establishes boundaries of retinoic acid signalling

EMBO J. 17 (24), 7361-7372 (1998)

99077803

2 (bases 1 to 1479)

Chen, J.L., Grunz, H., Panitz, F., Pieler, T. and Holleman, T.

Direct Submission

Submitted (05-APR-1998) Developmental Biochemistry, University of

Goettingen, Humboldtallee 23, Goettingen, Lower Saxony 37073,

Germany

FEATURES

source

CDS

1..1479

/organism="Xenopus laevis"

/db_xref="taxon:8355"

1..1479

/note="RACE"

/codon_start=1

/product="retinoic acid converting enzyme"

/protein_id="AAC25158.1"

/translation="MDLYTLTSLAQTALPLLLTLFAKLMEVYCLRRKDAACANPL

PGCTMKLPEFGTTLQORRRFLQYKRSQYRITKTLFGSPYRVATYGAENVROILM

GERKLVSYRMPASVRIITLACLSINLHDEHYTKVLAQASREALANVYDMOESEV

RCSVNLMDSGPCVLYPAIKRMFRIRAKRLGLCDPQRMDEQCELTLEAFEESSRN

LESLIPDVPSEGLYRGLRARNLHAQIEINIKELQREDEHKDALQLLIDYSRNG

EPINLQALKEATLFLGHTGTAASATSLTFLALHDKVLEKRVKLETLQGLSTKP

EEKKELSLVLOQIKYTSKIKETLRLSPVAGGFVALKTFVLNGYQIPKGMVYS

IAQTHGADLPDPTDKFNDFRLLPLPRDSRFGLPFGGVRCCIGKEFAKILLVF

VVELCNKCMELNLSPMATTSPIICPVNDLPAKRPSSSI"

```

BASE COUNT      350 a      422 c      384 g      323 t
ORIGIN
alignment_scores:
    Quality: 1646.50      Length: 494
    Ratio: 3.892      Gaps: 4
    Percent Similarity: 85.628      Percent Identity: 63.158
alignment_block:
US-09-668-482-2 x AF057566 ..
Align seg 1/1 to: AF057566 from: 1 to: 1479

1 MetGlyLeuTyrThrLeuMetValThrPheLeuGlyThrIleValLeuPr 17
  ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1 ATGATGCTGTACTCTGCTGCTACACGCGCTGTGTGACACCTGGACATGCC 50
17 oValLeuLeuPheLeuAlaValAlaValLysLeuTrpGluMetLeuMetIleA 34
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
51 ATGCTGCTCTCTCTCAACCGCTGCCAAACTCTGGGAAGTCTACTGCTGCA 100
34 rGArGValAspProAsnGlySerGlySerProLeuProProGlyThrMetGly 50
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
101 GGGCAAGAGATGCCGCTGCGCTACCCACTGCGCCCGGGCACCATGGGA 150
51 LeuProPheIleGlyGluThrLeuGluLeuIleLeuGluInArGArGlySph 67
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
151 CTGGCTTCTTCTGGAGAGACTCTGCAAAATGCTGTCTACAGAGCGCGAGTT 200
67 eLeuArGMetLysArGLeuLysTyrGlyCysIleTyrLysThrHisLeuP 84
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
201 CTTCCCAAGTCAAGGCAAGCAATACGCTGCATCTACAAGACGATCTGT 250
84 heGlyAsnProThrValArgValMetGlyValAspAsnValArgGluIle 100
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
251 TCGGCAACCCACAGGTCGCGCTACGCGGGGGGAGAACGTTGCAACATC 300
101 LeuLeuGlyGluHisLysLeuValSerValGlnTrpProAlaSerValAr 117
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
301 CTGATGGGGACACACAAAGTGGTGGTGGTGCACATGGCGGCGCTGTGCG 350
117 gTrpIleLeuGlySerAspThrLeuSerAsnValHisGlyValGlnHisL 134
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
351 CACCATCTCAAGGGGCGGCTGTCTGTCACACGTCACGACGACGACGACA 400
134 yAsnLysLysLysAlaIleMetArgAlaPheSerArgAspAlaLeuGlu 150
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
401 AGTACACCAAGAAAGTGATGCGCAAGCCTTCTCCCGGAAGCCCTGCT 450
151 HisTyrIleProValIleGlnGlnGluValLysSerAlaIleGlnGluTr 167
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
451 AATTAGGTGCGCGAGATGGAAAGAGGTGAGATGCTCCGTGAACCTGTG 500
167 pleuGlnLysAspSerCysValLeuValTyrProGluMetLysLysLeuM 184
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
501 GGTGGAGAGCGGCGCTGTGTGTGTGTGTATCCCGCATCAAAAGCAATGA 550
184 etPheArGtlleAlaMetArgIleLeuLeuGlyPheGluProGluGlnIle 200
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
551 TGTTCGATTCATGCAAGGCTCCTGCTGCTGATGCCCAACGCAATG 600
201 LysThrAspGluGlnGlu..LeuValGluAlaPheGluGluMetIleLys 216
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
601 GACCGGACAGAGAGAAAGCTGTGCGAAGCCTTGAGAGAAATGAGCG 650
216 sAsnLeuPheSerLeuProIleAspValProPheSerGlyLeuTyrArg 233
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
651 AATCTTTTCTCTTGGCTATGTGATGCGCGTTTACGCGGCTCTTACCGGG 700
233 LysLeuArgAlaArgAsnPheIleHisSerLysIleGluLysAsnIleArg 249
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
701 GTCTGGGGCGAGAAACCTAATTTCATGCCCCAAATGAGAGAAACATCAAG 750

```

```

250 LysLysIleGlnAspAspAsnGluLysGlnGluLysTyrLysAspAl 266
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
751 GAAAAGCTGCAA.....AGAGAACCATGAATCAATTGCAAGGATGC 791
266 AlenGlnLeuLeuIleGluAsnSerArgArgSerAspGluProPheSerL 283
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
792 CTGTGACGTCTGATTGACTATGATGCGGAGAAATGGCAGCCCATCACT 841
283 euGlnAlaMetLysGluAlaIleThrGluLeuLeuPheGlyGlyHisGlu 299
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
842 TGCAGCATGTGAAGATGCCGAAGTGAAGTCTGTTGGAGGCCACCGGA 891
300 ThrThrAlaSerThrAlaThrSerLeuValMetPheLeuGlyLeuAsnTr 316
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
892 ACCAGGCAAGTGTGCGCACGTCCTCACCTTTTCTTGGCGCTTACACA 941
316 rGluValAlaGlnLysValArgGluGluValGlnGluLysValGluMetG 333
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
942 AGAGTCTCTGAAAAAGTCCCGCAAGAGCTTGAACACAG.....G 982
333 LysMetTyrThr.....ProGlyLysGlyLeuSerMetGluLeu 345
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
983 GTCTCTGTCCACTAAACCCGAGAAAAAGAGAACTGACATGGAAGTA 1032
346 LeuAspGlnLeuLysTyrThrGlyCysValIleLysGluThrLeuArgIl 362
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1033 CTTCACACAGTTAAAGTATACAGCTGTGTCAATTAAAGGAGACTCTGCGCT 1082
362 eAsnProProValProGlyGlyPheArgValAlaLeuLysThrPheGlu 379
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1083 CAGCCACACAGTGGCGGCTGATTTCCGCTGCTCTCAAGACGTTTGTAT 1132
379 euAsnGlyTyrGlnIleProLysGlyTrpAsnValIleTyrSerIleCys 395
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1133 TAAATGGCTATCAAGATCCCAAGGGCTGGAATCTTATATACAGATTGCT 1182
396 AspThrHisAspValAlaAspValPheProAsnLysGluGluPheGlnPr 412
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1183 GACACCCACAGCTGAGGCCGATGTGTTCGCCGACGAGCAAAATTCACACC 1232
412 oGluArgPheMetSerLysGlyLeuGluAspGlySerArgPheAsnTyrI 429
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1233 TGACCGGTTTCTAAAGCCACTCCCAAGATTCCTCAAGGTTTGGCTTCA 1282
429 LepProPheGlyGlySerArgMetCysValGlyLysGluPheAlaLys 445
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1283 TCCCTTTCGAGGTGGAGTGAAGGTGCTCATCGCAAAAGAAATTTGCTAA 1332
446 ValLeuLeuLysIlePheLeuValGluLeuThrGlnHisCysAsnTrpIl 462
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1333 ATCCTTCTCAAGGTGTTCGTTGAGGAATTTGTTCGTATATTCGACAGTGA 1382
462 eLeuSerAsnGlyProProThrMetLysThrGlyProThrIleTyrProV 479
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1383 GCTCTGAAAGAGATCCCGCCCATGACAACTCCCGATATATCTGTCCTG 1432
479 AlAspAsnLeuProThrLysPheThrSerTyr 489
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1433 TGGACAACTACCCGCTAAATTCAAACCTTTC 1464
seq_name: gp_hcg21:AL358613
seq_documentation_block:
LOCUS      AL358613      166337 bp      DNA
DEFINITION Homo sapiens chromosome 10 clone RP11-348J12, *** SEQUENCING IN
PROGRESS ***, 14 unordered pieces.
ACCESSION      AL358613
VERSION      AL358613.8  GI:13660841
KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

REFERENCE 1 (bases 1 to 166337)
AUTHORS Plumb, B.
TITLE Direct Submission
JOURNAL Submitted (13-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone
requests: clonerequests@sanger.ac.uk
On Apr 17, 2001 this sequence version replaced gi:13625010.

COMMENT ----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
Project Information
Center project name: ba348j12

----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid, L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 161818 bases at least Q40
Consensus quality: 163255 bases at least Q30
Consensus quality: 163957 bases at least Q20
Insert size: 165037; sum-of-contigs
Insert size: 176845; 5.5% error; agarose-fp
Quality coverage: 5.93x in Q20 bases; sum-of-contigs Quality
coverage: 5.81x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 14 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 32409: contig of 32409 bp in length
* 32410 32509: gap of 100 bp
* 32510 52913: contig of 20404 bp in length
* 52914 53013: gap of 100 bp
* 53014 55817: contig of 2804 bp in length
* 55818 55917: gap of 100 bp
* 55918 74512: contig of 18595 bp in length
* 74513 74612: gap of 100 bp
* 74613 83833: contig of 9221 bp in length
* 83834 83933: gap of 100 bp
* 83934 87001: contig of 3068 bp in length
* 87002 87101: gap of 100 bp
* 87102 97488: contig of 10387 bp in length
* 97489 97588: gap of 100 bp
* 97589 100583: contig of 2995 bp in length
* 100584 100683: gap of 100 bp
* 100684 105295: contig of 4612 bp in length
* 105296 105395: gap of 100 bp
* 105396 129702: contig of 24307 bp in length
* 129703 129802: gap of 100 bp
* 129803 132462: contig of 2660 bp in length
* 132463 132562: gap of 100 bp
* 132563 137099: contig of 4537 bp in length
* 137100 137199: gap of 100 bp
* 137200 158409: contig of 21110 bp in length
* 158410 158409: gap of 100 bp
* 158410 166337: contig of 7928 bp in length.

FEATURES
SOURCE
1..166337
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-348J12"
/clone_lib="RPC1-11.2"
1..32409
/note="assembly_fragment:02593
fragment_chain:1
clone_end:17

misc_feature
1..32409
/note="assembly_fragment:02593
fragment_chain:1
clone_end:17

misc_feature 32510..52913
/note="assembly_fragment:01673
fragment_chain:1"
misc_feature 53014..55817
/note="assembly_fragment:01466
fragment_chain:2"
misc_feature 55918..74512
/note="assembly_fragment:01282
fragment_chain:2"
misc_feature 74613..83833
/note="assembly_fragment:02258
fragment_chain:2"
misc_feature 83934..87001
/note="assembly_fragment:00938
fragment_chain:2"
misc_feature 87102..97488
/note="assembly_fragment:01557
fragment_chain:2"
misc_feature 97589..100583
/note="assembly_fragment:00199
fragment_chain:2"
misc_feature 100684..105295
/note="assembly_fragment:01309"
105396..129702
/note="assembly_fragment:02028.0"
129803..132462
/note="assembly_fragment:02077"
132563..137099
/note="assembly_fragment:02005
fragment_chain:3"
137200..158309
/note="assembly_fragment:00282
fragment_chain:3"
158410..166337
/note="assembly_fragment:01837
fragment_chain:3
clone_end:SP6
vector_side:right"
BASE COUNT 46344 a 33577 c 34990 g 50115 t 1311 others
ORIGIN

alignment_scores:
Quality: 1279.00 Length: 1166
Ratio: 2.933 Gaps: 22
Percent Similarity: 37.393 Percent Identity: 29.931

alignment_block:
US-09-668-482-2 x AL358613 ..

Align seg 1/1 to: AL358613 from: 1 to: 166337

1 MetGlyLeuTYrThrLeuMetValThrPheLeuCyshTrileValLeuPr 17
|||||
54914 ATGGGCTCCCGCGCTGCGTCCAGTGGCTGCGACCTTCGCTGCGCC 54963
|||||
17 OValLeuDeuPheLeuAlaValLysLeuTrpGluMetLeuMetIleA 34
|||||
54964 GCTGCTGCTTCTCCTGCTGCGATCAAGCTCTGGACCTGACTGACGTA 55013
|||||
34 rGArGValAspProAsnCyArGSerProLeuProProGlyThrMetGly 50
|||||
55014 GCGGCGGAGCGAGCTGTGCTCCATTCATGCCCCCGGAGCTATG6GC 55063
|||||
55064 TTCCTCTTTGGGAAACTCTGCAGATGTACTGCAGAGTAAGGAGG 55113
|||||
63 63
55114 TGGGCGGAGAGCTGCTCCCGGAGCCGCGCGGCTGTGGGCTTCT 55163


```

56951 TCCTCTGGAGACACAGAACCGCCAGTCAGCCACATCTGCATC 57000
310 MetPheLeuGlyLeuAsnThrGluValValGlnLysValArgGlnGluVal 326
57001 ACTTACCTGGGGCTCTACCCACATGTTCTCCAGAAAGTGGCAGAAAGAGCT 57050
326 LglnGlnLysVal...GlnMetGlyMet..... 334
57051 GAAGAGTAAGTAGGAGAGACTGGGTCTGGGGGTCTCCTATTAGACTTAGG 57100
334 ..... 334
57101 AAATTACGCTGCTCCCTAGCCACTCCGAATTAAGTACGTGTCTGCTT 57150
334 ..... 334
57151 CATGAGATTTTGAAGTGCAGAGCCCGGCTGCGTGGGCCAGTGGGC 57200
334 ..... 334
57201 AGAATTACCTTTGTGAATTAACAGATGAGAGTCTTGAGCTGTGAGCCAC 57250
334 ..... 334
57251 TTGGCAGGCTTTGACCTTGATTCTACCCCTCCTCCACCTTTGCTGA 57300
334 ..... 334
57301 ACCGTGAGATTTTCAGATAGGTTCCACTTCTTGAATTGCTGTCCAA 57350
334 ..... 334
57351 GGGCATACATAGTGTGATAGCTGATTAGTGTGGGTGTGTGAGTGTG 57400
334 ..... 334
57401 TGGGGTGGGGGTTTTTTTGGGGGAACCTTGCTCTTACTGCTGAGCAG 57450
335 TyrThrPro.....GlyLys.Gly. 340
57451 CATTCCTCCGCGATTTGTAATAGATAGTGTGAGTGTGGCAGCAAAATGGCC 57500
340 ..... 340
57501 AATTAGCTGCTTTCCCTCCCTCCAGTCTCTCCATCATGAGGGCCCTTGG 57550
340 ..... 340
57551 GTTATGCTCCACTTAAGCCCTGTTTACGCTGCTGGCTGATTTTATG 57600
340 ..... 340
57601 GAGCACAATAACTGTTCACCTCTGATGACTGTTTGTATAGGGTTTAC 57650
341 ..... 348
57651 TTTGCAAGACATCAAGACAAACAGTGTGCAATGGAAATTTTGGACAA 57700
349 LeuLysTyrThrGlyCysValIleLysGlnThrLeuArgIleAsnProPr 365
57701 CTTAAATACATGGGTGTGTATTATTAAGAGACCCCTTCGACTGAATCCCC 57750
365 ovalProGlyGlyPheArgValAlaLeuLysThrPheGlnLeuAsn... 380
57751 AGTTCCAGAGGGGTTTGGGTCTCTGAAGACTTTTGAATTAAGTAA 57800
380 ..... 380
57801 GTTAAATTTCTCTTCTTCTCCCTTTTGTGTGTTAAAAACTCCTCTT 57850
380 ..... 380
57851 TCTTCCTATGCTGTGGCTGCAATTTCTATGCTTTGATTAATTGTTCTGC 57900

```

```

380 ..... 380
57901 CCTATATGAGATATGTTTACAGCAACCTGGATCCACTCTCTTTTAC 57950
380 ..... 380
57951 CTCTCCCTGCTTTTACATTAATTTCCCAAGATATCAGTACTGC 58000
380 ..... 380
58001 TTTTGTGTTGAAGTTAAATTCAGTTTGTACGCTTGGGGTTTCAT 58050
381 .....Gly. TyrGlnIleProLysGlyTyrPAsnValIleTyrSe 393
58051 AATCTTTTGCAGGATACCATGATTTCCCAAGGCTGGAATGTTATCTACAG 58100
393 rLecyAspThrHisAspValAlaAspValPheProAsnLysGlnGluP 410
58101 TATCTGTGATACATCATGATGTGGCAGAGATCTTCACCAACAGAGAAAT 58150
410 heGlnProGluArgPheMetSerLysGlyLeuGlnLysPglSerArgPhe 426
58151 TTAATCCTGACCATTCATCTGCTCCTCACCAGAGAGATGATCCAGGTTG 58200
427 AsnTyrIleProPheGlnGlyLysSerArgMetCysValGlnLysGluP 443
58201 AGCTTCATTCATTTGGAGAGGCTTAGAGCTGTGTAGCAAAAGATTT 58250
443 eAlaLysValLeuLeuLysIlePheLeuValGlnLeuThrGlnHisCysA 460
58251 TGCAAAATCTTCTCAAAATATTACGTGAGAGCTGGCAGGCAATGTG 58300
460 snrTPIleLeuSerAsnGlyProProThrMetLysThrGlyProThrIle 476
58301 ACTGGCAGCTTCTAAATGAGACCTCTACAAATGAAGAAACGACTCCACCGT 58350
477 TyrProValAspAsnLeuProThrLysPheThrSerTyr 489
58351 TATCCTGTGACATCTCCCTCGCAAGATTCACCCCATTTG 58389

seq_name: gb_pf5:AF252297
seq_documentation_block: 4445 bp mRNA PRI 14-JUN-2000
LOCUS AF252297 Homo sapiens cytochrome P450 retinoid metabolizing protein
DEFINITION P450RA1-2 mRNA, complete cds.
ACCESSION AF252297
VERSION AF252297.1 GI:8515440
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 4445)
AUTHORS White,J.A., Ramshaw,H., Taimi,M., Stangle,W., Zhang,A.,
Everingham,S., Creighton,S., Tam,S.-P., Jones,G. and Petkovich,M.
Identification of the human cytochrome P450, P450RA1-2, which is
predominantly expressed in the adult cerebellum and is responsible
for all-trans-retinoic acid metabolism
Proc. Natl. Acad. Sci. U.S.A. 97 (12), 6403-6408 (2000)
JOURNAL 20300913
MEDLINE 2 (bases 1 to 4445)
REFERENCE White,J.A., Ramshaw,H., Taimi,M., Stangle,W., Zhang,A.,
Everingham,S., Creighton,S., Tam,S.-P., Jones,G. and Petkovich,M.
Direct Submission
TITLE Submitted (04-APR-2000) Cytochrome Inc., 116 Barrie Street,
Bioscience Complex, Suite 2424, Kingston, ONT K7L 3N6, Canada
JOURNAL Location/Qualifiers
FEATURES
1..4445
source /organism="Homo sapiens"
/db_xref="taxon:9606"
1..1539
CDS

```

```

/codon_start=1
/product="cytochrome P450 retinoid metabolizing protein
P450RAI-2"
/protein_id="AAF76003.1"
/db_xref="GI:851341"
/translation="M.FEGLDLSVALATLAACLVSVTLTLLAVSQQLMWLRLMAATROKS
CLPIPKSGMGPPLIGETGHWLLOSGFSSRRKRYGVNFKHLGRPLIRYTGAEVY
RKILMGHEHLVSTEMPRSTYRMILGQVSNISIGDIHRNKRKYSKIFSHALESTYLPK
IOLVLODTLRAMSHPEALINVOEAKLFRRAIVRLGFSIPEEDLGHLEPVGOVF
DNFSLPVDLPSPSGYRGRIOARQILQKLEKIRKIOCTOGKDYLDLADLILSSKE
HKEMTMORIKDGTLELIPFAAYATASASTSIMOLIKRPTVLEKLRDELRAHGLIHS
GGCPEDGTLRIDTLISGLRDLVYKRWARLFPISGGRTTVQTEPDLDFQIPKQMSV
MSTRDHDAFVKNVDFDRFSQARSBDKDFHLPFGGVKTCOLGHLKLF
LVLAVELASTSRFELATFPFRLITLVPLHVDGLSVKFEGLDSNONEILPETEAML
SATV"

```

```

BASE COUNT      893 a      1265 c      1240 g      1047 t
ORIGIN

```

```

alignment_scores:
  Quality: 1034.50      Length: 494
  Ratio: 2.811          Gaps: 8
  Percent Similarity: 74.494      Percent Identity: 42.308

```

```

alignment_block:
US-09-668-482-2 x AAF252297

```

```

Align seg 1/1 to: AAF252297 from: 1 to: 4445

```

```

1 MetGlyLeuYrThrLeuMetValThrPheLeuCysThrIleValLeuPr 17
  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
16 TTGGATCTGGTGGCGCGTGGCCACCTCGCGCTGGCTGGCTGGCTGGCT 65
17 ovalLeuLeuPheLeuAlaValLeuThrIleValLeuMetIleVal 34
  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
66 GACGCTGCTGCTGGCGCTGGCGACACCTGGCGACGCTGGCGCGCGCG 115
34 rgrArValAspProAsnCysArgSerProLeuProGlyThrMetGly 50
  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
116 CCACCTCGCGACAGAGCTGCAAGCTGCCATCCCAAGGAGATCCATGGC 165
51 LeuProPheIleGlyIleuThrLeuGlnLeuIleuGlnArgGlySph 67
  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
166 TTCGCGCTCATCGGAGACCGCGCACCTGGCTGGCGAGGCTTGGCTG 215
67 eLeuArgMetLysArgGlnIleYrGlyCysIleYrLysThrHisLeu 84
  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
216 CCAGTCTGCGGAGGAGAGATGAGCAAGCTGTTCACAGACGATTTGT 265
84 heGlyAsnProThrValArgValMetGlyAlaAspAsnValArgGlnIle 100
  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
266 TGGGCGGCGCGCTGATACGCGTGAACCGCGCGAGAACGTGCGAAGATC 315
101 LeuLeuGlyGlnHisLysLeuValSerValGlnTrpProAlaSerVal 117
  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
316 CTCAATGGGCGAGACCAACCTCGTGAACCGAGTGGCTCGCACACCGG 365
117 gThrIleLeuGlnIleYrSerAspThrLeuSerAsnValHisGlyValHisL 134
  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
366 CATGTGCTGGGCGCCCAACAGGTGTCCAAATTCATTGGCGACATCCAC 415
134 ysaAnLysLysValAlaIleMetArgAlaPheSerArgAspAlaLeuGln 150
  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
416 GCAACAAAGCGCAGAGTCTTCCCAAGATCTTACACCAAGAGCCCTGGAG 465
151 HisTyrIleProValIleGlnGlnValLysSerAlaIleGlnGlnIle 167
  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
466 AGTAACTGCGCCCAAGATCCAGCTGTGATCCAGGACACACGCGCGCTG 515
167 pLeuGlnLysAspSerCysValLeuValIleProGlnMetLysLysLeu 184
  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
516 GAGCAGCCACCCGAGGCGCATCAAGGTGTACAGGAGGCGCAGAAGCTGA 565

```

```

184 ePheArgIleAlaMetArgIleLeuGlyPheGln...ProGlnGln 199
  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
566 CCTCCGCGCATGGCCATCCGCGTGTCTGCTGCTGCTGCTGCTGCTGAGAG 615
200 IleYrThrAspGlnGlnIleValGlnAlaPheGlnGlnMetIleYr 216
  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
616 .....GACCTGGCGACCTCTTGTAGGTCTTACACCAAGTGTGTGA 656
216 ssaLeuPheSerLeuProIleAspValProPheSerGlyLeuYrArg 233
  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
657 CATGTCTTCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 706
233 lYleuArgAlaArgAsnPheIleHisSerLysIleGlnGlnAsnIleArg 249
  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
707 GCATTCAGGCTGGCGAGATCTCCAGAACAGGCGTGGAGAACGATCCGG 756
250 LysLysIleGlnAspAspAsnGlnAsnGlnGlnIleYrLysAspAl 266
  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
757 GAGAACCTGCGAGTGCACACAGGCAAGGAC.....TACTTGGAGCC 797
266 aleuGlnLeuLeuIleGlnAsnSerArgArgSerAspGluProPheSerL 283
  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
798 CCGGACCTCTCTCATGAGAGCAGCAAGGAGCAGCGAGAGATGACCA 847
283 euGlnAlaMetLysGlnAlaIleThrGlnLeuLeuPheGlyGlnHisGln 299
  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
848 TGCAGAGCTGAAAGACGGGACCTGGAGCTGATCTTGGCGCTATGCC 897
300 ThrThrAlaSerThrAlaThrSerLeuValMetPheLeuGlyLeuAsnTh 316
  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
898 ACCACGCGCAGCGCCGACACCTCCTCATCATGACGCTGCTCAACACCC 947
316 rglValValGlnLysValArgGlnGlnValGlnGlnLysValGlnMet 333
  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
948 CACTGTGCTGAGAACCTCGGAGTATGCTGGCGGCTCAT.....G 988
333 lYMetYr.....ThrProGlyLysGly...LeuSerMetGlu 344
  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
989 GCATCTGCAACAGTGGCGGCTGCCCTGGAGGCGCACATCGCGCTGGAC 1038
345 LeuLeuAspGlnLeuLysYrThrGlyCysValIleLysGlnThrLeuArg 361
  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1039 ACGCTGATGCGGCTGCGCTACCTGCGACTGCTGCTCATCAAGAGTATGCC 1088
361 gIleAsnProProValProGlyGlyPheArgValAlaLeuLysThrPheG 378
  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1089 CCGTTCACGCGCCATTTCCGCGGCTACCGCACCTGTGCTGACAGCTTGC 1138
378 lueuAsnGlyYrGlnIleProLysGlyTrpAsnValIleYrSerIle 394
  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1139 AGCTTATGCTTCCAGATCCCAAGGCTGAGAGTCTCATGTATAGATC 1188
395 CysAspThrHisAspValAlaAspValPheProAsnLysGlnGlnPheG 411
  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1189 CGGAGATCCCATGACACAGCGCGCGGTTCAAAGACGTGAAGCTGTGCA 1238
411 nProGlnArgPheMetSerLysLysLeuGlnLysPhe...GlySerArgPhe 427
  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1239 CCGCGATCGCTTACCGACGCGGAGCGAGGACGACGATGCGCGCTTCC 1288
427 snTyrIleProPheGlyGlyGlySerArgMetCysValGlyLysGlnPhe 443
  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1289 ATTACGTCGCGTTCGCTGGCGGCTGCTGCGAGCTGCTGCGCAAGACCTG 1338
444 AlaLysValLeuLeuLysIlePheLeuValGlnLeuThrGlnHisCysAs 460
  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1339 GCCAACCTGTTCTGAAGGTGCGGCGGTGAGCTGAGCTGAGACCAAGCCG 1388
460 nTrpIleLeuSerAsnGlyPro...ProThrMetLysThrGlyProThrI 476
  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1389 CTTTGAAGCTGCTACACGAGACCTTCCCGCATCACTGTGCTCCCGCTCC 1438
476 lYrProValAlaAspAsnLeuProThrLysPhe 486

```

```

:::|||||:::| | :::|||||
1439 TGCACCCCGTGATGGCTCAGCGTCAAGTTC 1470

```

seq_name: gb_ba3:SYCSLE

seq_documentation_block:	
LOCUS	SYCSLLE_113064 bp DNA BCT 13-FEB-1999
DEFINITION	Synechocystis sp. PCC6803 complete genome, 22/27, 2755703-2868766.
ACCESSION	D64003 AB001339
VERSION	D64003.1 GI:1001200
KEYWORDS	TRNA-fMet; 2-ketoadid dehydrogenase (malate dehydrogenase, lactate

KEYWORDS
tRNA-met; 2-ketoadoic dehydrogenase (malic dehydrogenase); 30S ribosomal protein S18; 50S ribosomal protein L33; 6-aminohexanoate-cyclic-dimer hydrolase; 7-keto-8-aminoheptanoic acid synthetase; KdsB; Mg-protoporphyrin IX monomethyl ester oxidative cyclase; 66 kD subunit; Moxr protein; NADH dehydrogenase subunit 5; Moxr; Pled; acetyl-coenzyme A synthetase; acyl-CoA desaturase 1; alanine racemase; aspartate 1-decarboxylase; cyanate lyase; cytochrome P450; dtdp-glucose 4,6-dehydratase; elongation factor EF-G; endo-1,4-beta-glucanase; ferredoxin-nitrite reductase; ferredoxin-thioredoxin reductase, catalytic chain; ferredoxinase; flavoprotein; low affinity sulfate transporter; methionine aminopeptidase; methyltransferase; molybdenum cofactor biosynthesis protein A; molybdenum cofactor biosynthesis protein C; molybdopterin (MP) converting factor, subunit 2; molybdopterin biosynthesis Moen; oligopeptide transport system permease protein; peptidyl-tRNA hydrolase; phosphoribosyl formylglycinamide cyclo-ligase; photosystem II P680 chlorophyll A apoprotein; polyA polymerase; endonuclease; replicative DNA helicase; succinate dehydrogenase iron-sulfur protein; tRNA-Cys; tRNA-met(exon2); tRNA-met(exon2); Yjc34.
Synechocystis sp. (strain:PCC6803) DNA.

REFERENCE
Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
1 (bases 1 to 113064)
Kaneke, T., Tanaka, A., Sato, S., Kotani, H., Sazuka, T., Miyajima, N.,
Sugita, M. and Tabata, S.
TITLE
Sequence analysis of the genome of the unicellular cyanobacterium
Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
region from map positions 64t to 92t of the genome
DNA Res. 2 (4), 153-166 (1995)
JOURNAL

/transaltion="500ONYMMKRRPWCPRSGN1SLBDGXYLDDPMTKGRKGNKJDL
 /leiatipclivlgeBgvGKthLEKtOFTTEETOIKVGLMPTFMKJLKDILR
 DKTF50WLEGH1H1PLD5LDBELWVKSALGYDDEKRRKYS31NRH1LRTOR
 TLFP51LEBALKAME5SNFATIELAPURDYO IAVABEG5SDPTLAK1NORDIV
 PLAK1P51SGFLMNTYRRHNS5NPPEOKRRE5TECEKCL5EVC5CHAKLEKNGD
 IDORL1VARI5AATV1FANRFAW1DVEGQ5VAEDVPQK1CF5E5SNRE5BE1TQ
 EY1EVL1DGL5E5SG1LRMGWMAHQ1AE5FFPA1Q1ND1K5L5ITL5L5AD5P5T
 KL1POLHE7NAW1ASMRDNP5FOE1K1TDP5V1LQOT1D1P5ELRAT1VAN1LMO1EE5E
 KL1DPR5C5VW1YAK1K5H5GABOL1KY1D5SS5K5R5TR1A1D1A1C5N1S5G1O1
 AK1TL1P5OP1S1RYNARAL1S5IGD5T5L5K1K5AL1V1P5ED5ODEL5KAY1KALAM1
 PROUL5E1P5K1N1OL5MART5F5E5L5V5P5L5Q5ED5V1A1D5Q5O5S1L5G5
 HPEE1EGD1IL1KAEH5VNP5GL5E5G7FAL1Q5H5E5K1L5H5P55L5Q5A5F5O5S1L5G5
 1N5OKR5AL1E5E5V5R5D5E5Y5L55SM5P5E1S1S1D5V5M5E5K1O5D5NE5K1O5
 RAAVAK1I5E5FN5P5R5D5TQ1S5E1KATONNS1L5D5V5Y5F5T1E1D5E5K1O5D5H5L5
 KMN5NS1D5E5D5P5P5K5E5V5D5E5O5L5E5S1D5AM5R1N5R5E5M1L5P5Q5K5R5GH5M1K1L5
 L7O1P5W5K5AB5O5A5T5K1L5G1A1O1Y1O5G1O5N1D1Y5M1N5R5E5M1L5P5Q5K5A1L1L5
 EAP5R1D5N5T5D1M5K5R5M5P1L55V5N5F5D5N5P5K5R5C5V1A55H55P5DE1L5T5Y5L5
 L1DK5E5OD5N5N1Y5ID5LE5K5M5DE5K5L5N5E5K5D5Q5L5P5Y5K5L5E5K1EG5Q5A5O5A
 KDJ1AL5T1N5Y5N5E5S5D5E5R5O5K5AL5K1AL1T1L5D5D5SN5V1W5E5K1AD5V5G5E5A5E
 LA5SC5M5S1K1E5D5L5E5K1O5AD5V5AM5V5P5H5E5D5H5N5E5MA5V5P5R5ON1A5F5FN5
 N5S1L5O5R5D5G7T5E5AC5AE1L5R1O5E1P5H1M5K5E5Y1A5O5O5N5R5E5M5P5D5K1L5E5
 L1P5E5N5D1S1N5O5S1T5DK5R5E5K5E5D5E5P1K1N1N1D5S1T1NAP5T5S5V5T5N5N5Y5G5S5
 PE5OK5R5N5E1L5G5V1G1L5V1A5V5G1F5F5N5G1F5N5D5L5D5N5G5P5K1E5Q5AMP5K5N5

COMMENT	FEATURES
Potential protein coding regions were assigned on the basis of similarity search of the ORFs and Genemark analysis.	source Location/Qualifiers 1..113064 /organism="Synecocystis sp." /strain="PCC6803" /db_xref="taxon:1143" complement(1..3960)

CDS

```

/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="BAAL0445.1"
/db_xref="GI:1001205"
/translation="MSGKRLRSKSTOPSOLLTANAVTYSOGVDALTIVASAEAG
SGKGLHYPTKEALVAGVVOALDRFPETTLHOELANPADPDGHWRAVIRASALD
DOENEPHLNLANFTNPDELLAPMRHHWEOCHGIVASGLDPNAVATVIRLAMDGLWL
THLHGPAEPDEPLRSQVATATALKAQL"
8165 .9550
/note="ORF_ID:slr0896"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="BAAL0446.1"
/db_xref="GI:1001206"
/translation="WTDISMRFGMOAREFLKLAVPLAGAQMAQAATGEVDVVMGM
LGQDVLGAGGLATLPMAPMNTGVGLISGVPAEAVAGAGOTRIGQLFROGLMIVL
LISIGMGEITHLNCGVMMSQADPTILISQSYLVNVMAGFPAIARFATLRGCTIIS
QARPMILVIAATLFTNLINGNGLGGKMGKPGALITGTGLATASISAHIMELSILVYHL
MKRPVSLSDSLHLHKRLIIOOLLIWGWGPGALITGAIVLEGLVITTSFPMKALDPLL
AAHOVSQVTVLPEVPLAMSYATNVRYGOMFQGOHMPQIRQALVSTIGALVEMTLTA
GIALAYRQQLITGILTDLPANDEALNVGISIMKIAAGVLYLDGLORTNGVYIOGLGQ
DTRIDPWAGITLAWYIGILTASYLILGFYTPISGAVWIGTYIGLAASIAVIMRFQKVA
VRKSLPVARS"
9852 .13064

```

```

gene      13423..14931
/gene="nirA"
CDS       13423..14931
/gene="nirA"
/note="ORF_ID:slr0838"
/codon_start=1
/transl_table=1
/product="Ferredoxin-nitrite reductase"
/protein_id="BAA10448.1"

```

```

alignment_scores:
  Quality: 693.50      Length: 487
  Ratio: 2.188         Gaps: 13
Percent Similarity: 65.092      Percent Identity: 33.676

alignment block:
US-09-666-482-2 x SYCSLIE ..

Align seg 1/1 to: SYCSLIE from: 1 to: 113064

```

10 PheLeuCysThrIleValLeuProValLeuPheLeuAlaAlaValLys 26

```

62118 TTGGTAGTACTAATATCGAACAATCAACTATTTCTTCCATGATGA.. 62165
26 sLeuTrpGluMetLeuMetIleArgValAspPro...AsnGysArgS 42
62166 .....ACTTCTCTACCAAAATTAAAT 62187
42 er...ProLeuProProGlyThrMetGlyLeuProPheIleGlyIuThr 57
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
62188 CTTTACCACTATCCCCCGGGAGACTTTGTTCACCGTGTTAGAGAACT 62237
58 LeuGluLeuIleLeuInlArgArgLysPheLeuArgMetLysArgGlnLys 74
|||::: ::::|::: |||::: |||::: |||::: |||::: |||:::
62238 CTCAAA...TTTCTTATATAGTGGCGACTCTGGTAAAAAGCGGCACACAGA 62284
74 sTryGlyCysIleTryLysThrHisLeuPheGlyAsnProThrValArgV 91
:::|::: |||||::: |||||::: |||||::: |||||::: |||||:::
62285 ATTTGGCCCAATTTTAAACCCGTTTATTTTGGCAAAAATGTCATTTTAA 62334
91 aMetCysValAspAsnValArgGlnIleLeuLeuGlyGlnHisLeu 107
:: |||||::: |||::: |||::: |||::: |||::: |||:::
62335 TTTCTGGGGCATTAGCGAATTAATTTTATTCACCAAGAGCAGGAAACT 62384
108 ValSerValGlnTrpProIleSerValArgThrIleLeuGlySerAspTh 124
:::|::: |||||::: |||||::: |||||::: |||||::: |||||:::
62385 TTTCAGCAACTTGGCCCTCAGCCACCCAAATTTTGTTAGGACCTAAATGC 62434

```

[illegible]

```

62850 GAA.....GAGAGCGATTGGCATATTATTGGCCGCCG 62854
274 TATGATSerAspIuProPheSerLeuGlnAlaMetLeuGlyAlaAlaIat 291
      :::::|||||:::|||||:::|||||:::|||||:::
62855 AGATGATTAATACCAACCGTTATCTTGGCGGAGCGTAAGACGCAAAATT 62904
291 hrgIuLeuPheGlyGlyHisGluIuThrhrhIaSerTrpAlaTrnSer 307
      |||||:::|||||:::|||||:::|||||:::|||||:::
62905 TATTCGATATCTTTGGCGGACATGAACCTTTAACTTCGCGATTATCTTCT 62954

```

[illegible]


```

290 lathrgluLeuLeuPheGlyLysHisgluThrAlaSerThrAlaThr 306
    :: ||||| ::|||::|||::|||::|||
862 TGGTGGCTTACTGTGCGCGGTATGAACAACCTCCAGCATCATGCT 911
    ::|||::|||::|||::|||::|||
307 SerLeuAlaMetPheLeuGlyLeuAsnThrGluValAlaGlnLysValAr 323
    ||| ||||| ::| ::|||::|||::|||
912 CTCGCCCTCAATTCTCCACCGAGCTCTTACTCTCTGCTCAACCA 961
    ::|||::|||::|||::|||::|||
323 ggluGluValGlnGlnLysValAlaGluMetGlyMetGlyThrProGlyLysG 340
    ::|||::|||::|||::|||::|||
962 GGAAGAG...CATGAAAAGATTAG...GCATATAGAGTATTCGTATA 1005
    ::|||::|||::|||::|||::|||
340 LysSerMetGluLeuLeuAspGlnLeuLysTyrThrGlyCysValIle 356
    ::|||::|||::|||::|||::|||
1006 GCTTTCATGAGTACGATTACAAATGCAATGCCATTCACAAATGTGGT 1055
    ::|||::|||::|||::|||::|||
357 LysGluThrLeuArgIleAsnProProValProGlyIlePheArgValAl 373
    ::|||::|||::|||::|||::|||
1056 AATGAGAGCTACGAGTGGCTAACATCATCGCGGTGTTTCACAGCGTGC 1105
    ::|||::|||::|||::|||::|||
373 aLeuLysThrPheGluLeuAsnGlyTyrGlnIleProLysGlyTrpAsn 390
    ::|||::|||::|||::|||::|||
1106 AATGACGAGATTGAGATCAAGGTTTAAATTCCAAAGAGGTGGAAG 1155
    ::|||::|||::|||::|||::|||
390 aLleLysSerIleCysAspThrHisAspValAlaAspValPheProAsn 406
    ::|||::|||::|||::|||::|||
1156 TATTCATCATGTTTAGACGCTTACATTAACCAACCACTTCAAAGAT 1205
    ::|||::|||::|||::|||::|||
407 LysGluGluPheGlnProGluArgPheMetSerLysGlyLeuGluAspG 423
    ::|||::|||::|||::|||::|||
1206 GCTCGCACTTCAACCTTGGAGTGGCAGACAACTCGGTACAGCAGG 1255
    ::|||::|||::|||::|||::|||
423 YSerArgPheAsnTyrIleProPheGlyGlySerArgMetCysValG 440
    ::|||::|||::|||::|||::|||
1256 CCTTTCATATGTGTTCACACCGTTTGGTGAGGCGCAAGCTATGTCGG 1305
    ::|||::|||::|||::|||::|||
440 LysGluPheAlaLysValLeuLeuLysIlePheLeuValGluLeuThr 456
    ||| ||||| ::|||::|||::|||::|||
1306 GTTACAGAGCTGGTGGTGGTCACTCTCTTCTTCCACCGCTTACGT 1355
    ::|||::|||::|||::|||::|||
457 GlnHisCysAsnTrpIleLeuSerAsnGlyProProThrMetLysThrG 473
    ::|||::|||::|||::|||::|||
1356 ACAGGCTTCAGTTGGGT.....CTTCACAGCAAGCAACAA 1390
    ::|||::|||::|||::|||::|||
473 YProThrIleTyrProValAspAsnLeuProThrLysPheThrSerTyr 490
    ::|||::|||::|||::|||::|||
1391 GCTGATTCTTCCAACTACAAAGACGAGAAAGCTACCGATCTTCG 1440
    ::|||::|||::|||::|||::|||
490 aLArg 491
    ::|||::|||::|||::|||::|||
1441 TGAG 1445
    ::|||::|||::|||::|||::|||
seq_name: gb_p12:ATCYP450R
seq_documentation_block:
LOCUS ATCYP450R 1608 bp mRNA PLN 13-MAY-1996
DEFINITION A. thaliana mRNA for cytochrome P450.
ACCESSION X87367
VERSION X87367.1 GI:853718
KEYWORDS Cytochrome P450.
SOURCE
ORGANISM Arabidopsis thaliana
Eukaryote; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 1608)
AUTHORS
Altmann, T., Redei, G.P., Nagy, F., Schell, J. and Koncz, C.
Brassinosteroids rescue the deficiency of CYP90, a cytochrome P450,
controlling cell elongation and de-etiolation in Arabidopsis
JOURNAL
Cell 85 (2), 171-182 (1996)
MEDLINE
96200769
REFERENCE
2 (bases 1 to 1608)

```

```

AUTHORS Szekeres, M.
TITLE Direct Submission
JOURNAL Submitted (22-MAY-1995) M. Szekeres, Inst. of Plant Biology,
Biological Research Center, Hungarian Academy of Sciences, PO Box
521, 6701 Szeged, HUNGARY
FEATURES
source
1..1608
/organism="Arabidopsis thaliana"
/cultivar="columbia"
/db_xref="taxon:3702"
/chromosome="5"
/clone_lib="lambda gt10"
/clone="C204"
/map="14.3 CM"
48..1466
/gene="CPD"
48..1466
/gene="CPD"
phenotype="loss of gene function results in dwarf
/feature="loss of gene function results in dwarf
/note="cytochrome P450"
/codon_start=1
/product="CYP90 protein"
/protein_id="CAA60793.1"
/db_xref="GI:853719"
/db_xref="SWISS-PROT:Q42569"
/translation="MAFTRFLLSLAAGFLLLRRTRYRMGLPPGSLGLPGET
FOLIAVKTENPEPFLDERVARYGVFMTHLGEPTISADETRNFVLQNGKLEEC
SYPASTICMLNKHSLMLKMSGLHKRMHSITMSFANSIIKDLMLDIDRLVFNLDWS
SRVILMEBAKITEFLVYKQLMSFDGEMSESLREYLLVTEGFSLPLPLEFTYR
KAIQARRKVAELTVVMKRRREEDEGKERKMDMLAIIAADGSDSEIYDFVAIL
VAGETSTIMTLAVKPLTEPLAIAOLKEHEKIRANKSDSYSLPMSEKMPPTOC
VWNETLRVANTIGCYERRAMDVLEIKYKIRPGMKVFSFRVHIDPLNFKARFNP
KRMQSNSTGTSNFTPPGGPRLCPGIELARVALSVLHRLVTFGSWVPAEDKLV
FFPTTKRTPPIEFVKRDFAF"
BASE COUNT 382 a 374 c 374 g 478 t
ORIGIN
alignment_scores:
Quality: 552.50 Length: 485
Ratio: 1.811 Gaps: 14
Percent Similarity: 62.887 Percent Identity: 29.897
alignment_block:
US-09-668-482-2 x ATCYP450R ..
Align seg 1/1 to: ATCYP450R from: 1 to: 1608
19 LeuLeuPheLeuAlaAlaValLysLeuTrpGluMetLeuMetIleArg 35
|||||::|||::|||::|||::|||
66 CTCCTCCTCCTCTCTTCATCCATCGCGCGGCTTCTCTCTCTCTCTCT 115
|||::|||::|||::|||::|||
35 gValAspProAsnCysArgSerProLeuProProGlyThrMetGlyLeuP 52
|||||::|||::|||::|||::|||
116 TACACCTTACCGT...CGGATGGGTCTGCTCCGGAAGCTGTGCTCC 162
|||::|||::|||::|||::|||
52 ropheLleGlyGluThrLeuGlnLeuIle.....LeuGln 63
|||||::|||::|||::|||::|||
163 CTCTGATGAGAGAGACTTTTCAGCTGATCGGAGCTTACAAACAGAGAC 212
|||::|||::|||::|||::|||
64 ArgArgLysPheLeuArgMetLysArgGlnLysTyrGlyCysIleTyrI 80
|||::|||::|||::|||::|||
213 CCGGACCTTTCATCGACGAGAGAGTACCGCGGTACGTTGGTTCAT 262
|||::|||::|||::|||::|||
80 sPThrIleuPheGlyAsnProThrValArgAlaMetGlyAlaAspAsn 97
|||||::|||::|||::|||::|||
263 GAGCAGCTTTTGGTGAACGACGATTTTTCAGCTGACCGGAAACGA 312
|||::|||::|||::|||::|||
97 aLArgGlnIleLeuGluGlyGlnHisLysLeuValSerValGlnTrpPro 113
|||::|||::|||::|||::|||
313 ACCGGTTTCTTCTTCAGACGAGGAGACCTTTTGAGTGTCTTATCT 362

```


Percent Similarity: 62.611 Percent Identity: 30.752

alignment block:

US-09-668-482-2 x AF318211

Align seg 1/1 to: AF318211 from: 1 to: 1494

```

17 ProValLeuLeuPheLeuAlaValLysLeuTrpGluMetLeu11 33
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
61 CCTGCTATCTTCTCCCTACCCGACACCAATTCCTGCAATTCCTCT 110
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
33 eArArValAspProAsnGlySer.....ProLeuProGlyTr 48
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
111 CTTCCTCTCCGTTACATCCGATCCCTCTTAACTCCCTCGAA 160
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
48 hmeGlyLeuProPheLeuGlyTrpLeuGlnLeuLeuGlnArg 64
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
161 AGTTAGGTTTCTCTCATCGGGAGACCAATTAATTCGCGACTC 210
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
65 Arg.....LysPheLeuArgMetLysArgGlnLysTrpG 76
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
211 CGATCAGAAACCTCAAAAGTTTGTATGATGATGAAGAAATTCG 260
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
76 yCysIleTrpThrHisLeuPheGlyAsnProThrValArgValMetG 93
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
261 TCCTGTTACATGACTTCCCTAATGGGCAATCCACAGTTGTTACTGCG 310
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
93 lYAlAspAsnValArgGlnLeuLeuGlyGlnHisLysLeuValSer 109
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
311 GGCCCTGCGGAAACAATTAAGTTCTTCGAAACGAGACCACTGTTAG 360
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
110 ValGlnTrpProAlaSerValArgThrLeuGlySerAspThrLeuSe 126
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
361 ATGGAAGGCCCAAGCTCTTCATGAACATGATGGGAGAGTTCCATTTG 410
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
126 fAsnValHisGlyValGlnHisLysAsnLysLysAlaIleMetArg 143
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
411 TCGTAAAGAGCGCGAGATCATCGATCTTACGCACCTGCTGCTGCT 460
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
143 lApheSerArgAspAlaLeuGlnHisTrpIleProValIleGlnGln 159
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
461 TTTTGGCGCTCAAGCTTACAAAATTAATCGGGAGATGAGTTCCAGA 510
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
160 ValLysSerAlaIleGlnGlu...TriPleuGlnLysAspSerCysVal 175
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
511 ATAGGACACCATTCATGAATAAATGGAAGGTAAAGATGAA...GTGA 557
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
175 vAlTrpProGlnMetLysLysLeuMetPheArgIleAlaMetArgIle 192
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
558 GGTGCTTCTTGGTGAAGAGGCTTATCTCTCCATGTCAGACACCTGT 607
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
192 euleGlyPheGlnProGlnIleLysTrpAspGlnGlnLeuVal 208
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
608 TTTTCGATGAATGATGACAC.....CAACAGAACCACTTCAT 648
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
209 GluAlaPheGlnGlnMetIleLysAsnLeuPheSerLeuProIleAsp 225
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
649 CATCTTCTGGAACATATCTTGTGGGAAGTTGTACGCTCGGACTT 698
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
225 lProPheSerGlyLeuTrpArgGlyLeuArgAlaArg.....A 238
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
699 TCCAGAGACCTGTTATCTGAAGGCTTCAGCGCGCTGAAGCTTGAT 748
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
238 snPheIleHisSerLysIleGlu.....GluAsnIleArgLysLys 251
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
749 AAATCTCTCTCTCTAATAAACCCAGACAGAGAGATCTCGTTACAG 798
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
252 lIleGlnAspAspAsnGlnGlnGlnLysTrpLysAspAlaLeuG 268
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
799 ATAGCTTCTGATGAT.....CAAGTCTACTCTC 827
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
268 nleuLeuIleGlnAsnSerArgArgSerAspGlnProPheSerLeuG 285
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

828 GGTGTTCTTCACCTTCAGATGAAAAAGGAACCTGACGACGACG 877
285 lmetLysGlnAlaIleArgLeuLeuPheGlyGlyHisGlnTrp 301
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
878 GGAATCTGACCAACTTCTGCTATGTTTCATGCTTCATATGACAC 927
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
302 AlaSerThrAlaThrSerLeuValMetPheLeuGlyLeuAsnThr 318
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
928 GTTGACCAATGCGCTTGATATTAACTTATCTTACCTCAATCCGA 977
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
318 lValGlnLysValArgGlnGluValGlnGlnLysValGln... 334
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
978 CCATGAAAAAGTA.....TTTCAAGACGAGTTGGAATATTCGA 1018
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
334 eTrpThrProGlyLysGlyLeuSerMetGlnLeuLeuAspGln 350
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1019 ATAAAAAGAGGAGGAAAGAAATCAGTTGAGAGATTTGAATCTA 1068
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
351 TrpThrGlyCysValIleLysGlnTrpLeuArgIleAsnProVal 367
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1069 TATACATGGCAGACATTCAGAGATCAGATACGATACCCACACT 1118
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
367 oglyLysPheArgValAlaLeuLysThrPheGlnLeuAsnGly 384
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1119 TCGAATATTTCGTAAGCTATCATCTGATTCATATGATGAGTAT 1168
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
384 lProLysGlyTrpAsnValIleTrpSerLeuGlyAspThrHis 400
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1169 TTCAAAAGATGAGGAGGTTTATGTTACCTTATCTACATCTGAGA 1218
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
401 AlaAspValPheProAsnLysGlnGlnPheGlnProGlnArgPhe 417
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1219 GAGAGTACTTCCCTGAGCCTGAAGATTCAGCCTTCAAGATTGAGA 1268
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
417 rLysGlyLeuGlnAspGlySerArgPheAsnTrpIleProPhe 434
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1269 TGAAGC...AGGCATGTGATCTCTTACACATATGATGATTTG 1315
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
434 lYserArgMetCysValGlyLysGlnPheAlaLysValLeuLys 450
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1316 GCGTGGCAGCATGTCAGGATGGGATTTCAAAAGATGAAATATTA 1365
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
451 PheLeu 452
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1366 TTTGTC 1371
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

seq_name: gb_p12:AF279252

seq_documentation_block: 1649 bp mRNA 30-JUL-2000
LOCUS AF279252 DEFINITION Vigna radiata cytochrome P450 (Cyp1P) mRNA, complete cds; nuclear
gene for chloroplast product.
ACCESSION AF279252 VERSION AF279252.1 GI:9587210
KEYWORDS mung bean.
SOURCE Vigna radiata
ORGANISM Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids 1;
Fabales; Fabaceae; Papilionoideae; Vigna.
REFERENCE 1 (bases 1 to 1649)
AUTHORS Yang,M.T. and Chen,Y.M.
TITLE Cloning and sequencing of a Vigna radiata cDNA encoding cytochrome
P450
JOURNAL Unpublished
2 (bases 1 to 1649)
AUTHORS Yang,M.T. and Chen,Y.M.
TITLE Direct Submission
JOURNAL Submitted (16-JUN-2000) Botany, National Taiwan University, No. 1
Sec. 4 Roosevelt Road, Taipei, Taiwan
FEATURES
1.1649
/organism="Vigna radiata"

```


[illegible]

```

301 rAlaserThrAlaThrSerLeuValMetPheLeuGly..... 313
68260 CTCCTACGCCCATGTCATGATCTTCACAGATGTAGCTCAGACCCAACTT 68309
314 .....LeuAnThrGluValVal 319
68310 CTCGAGACACTCTCTCAGGTAACTCAATCACTCAATATATGAGATGTC 68359
320 .....GlnLys..... 321
68360 TCTTCCTATCAGCAAAATATATATCTCACTCACTCAGTTTCTCTGAT 68409
322 ValArgGluGluValGlnGluValGluValMetGlyMetThrProG 338
68410 CATATCAGACATGCTCTCAATATAAAGCAACAAGG.....GAG 68450
338 lYlYsGlyLeuSerMetGluLeuLeuAspGlnLeuLysThrThrGlyCys 354
68451 GAGAAATATCTTACAGTGAAGATGAGAAAGATGAACTATAGTGGCA 68500
355 ValIleLysGluThrLeuArgIleAsnProProValProGlyLysPheAr 371
68501 GTTGTCTGTAACAACATGATATCCCTCAATCTCGGTTCTTCTAG 68550
371 gValAlaLeuLysThrPheGluLeuAsnGlyTyrGlnIleProLysGlyT 388
68551 AAACACATGCTGATATTGATGAGGATACAAATCCCAAGGCT 68600
388 rPasnVal..... 390
68601 GGAAGCTTAAACATCACTCTTAAGTCTTCACTTCTAATCAAA 68650
391 .....Ile-TyrSerIleCys..... 395
68651 CTCATTTCTTGAAGAAATACTAATAAATCAACCTCTCTTCTTT 68700
396 .....AspThrHisAspValAlaAspVal 403
68701 GAATCCAGATACTCTGCACCACTTATGAACTCACTTAATATCCGTAATA 68750
404 PheProAsnLysGluGluPheGlnProGluArgPheMetSerLysGlyLe 420
68751 TTTCAAGATCCATGATGATTCACCAAGATTT...GATAGCCCAAT 68797
420 uGluAspGlySerArgPheAsnTyrIleProPheGlyLysGlySerArgM 437
68798 ACAG.....GCATATACATATCTACCATTTGAGAGTGGACCAAGGC 68838
437 etCysValGlyLysGluPheAlaLysValLeuLeuLysIlePheLeuVal 453
68839 TTTGTCAGAGTCACCACTAGCCAAATACAGATCTCGTTTTCATGCAT 68888
454 GluLeuThrGlnHisCysAsnThrIleLeuSerAsnGlyProProThiMe 470
68889 TTTGTGTGACAGGTTTGTAGCTGCTTATGCTACCAAGATGAGACAT 68938
470 lYsThrGlyProThrIleTyrProValAspAsnLeuProThrLysPheT 487
68939 TAGCATGATCTCTCCATTTCCGCTCCCTTGGATGCCAAATCAAAATTT 68988
487 hrSerTyrVal 490
68989 CTCCCAAGTG 68999
seq_name: gb_p14:SL054770
seq_documentation_block:
LOCUS SL054770 1395 bp mRNA PLN 18-OCT-1996
DEFINITION Solanum lycopersicum cytochrome P450 homolog (Dwarf) mRNA, complete cds.
ACCESSION U54770
VERSION U54770.1 GI:1421740

```

```

KEYWORDS
SOURCE
ORGANISM
tomato.
Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliopsida; eudicotyledons; core eudicots;
Asterales; eunasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
REFERENCE
1 (bases 1 to 1395)
Bishop,G.J., Harrison,K. and Jones,J.D.
The tomato Dwarf gene is isolated by heterologous transposon tagging
encodes the first member of a new cytochrome P450 family
96266705
JOURNAL
MEDLINE
2 (bases 1 to 1395)
Plant Cell 8 (6), 959-969 (1996)
AUTHORS
Bishop,G.
TITLE
Direct Submission
JOURNAL
Submitted (11-APR-1996) Gerard Bishop, Sainsbury Laboratory, John
Innes Centre, Colney Lane, Norwich, Norfolk, NR4 7UH, UK
FEATURES
source
1..1395
/organism="Lycopersicon esculentum"
/strain="GCR758"
/db_xref="taxon:4081"
1..1395
/gene="dwarf"
1..1395
/gene="dwarf"
/codon_start=1
/product="cytochrome P450 homolog"
/protein_id="AAB17070.1"
/db_xref="GI:1421741"
/translation="MAPFLIFLSFPGICFTALLRMNOVKYNOKNLPPGTGMPF
GETTEFLGSPFKNORARSGSPFKSIIIECTIVMSDSLNTIITNNAKGLPYX
FOSMDILGKNIAVNGSAHKYRGALSLISPTMRDOLRIDEFKSHLTNRDN
KVIDQETKNMAFLSLKQJAGLESLSLAEFSEFNLTGLSLPINLPTNVR
GEOARKIIVNLRILIERASKKEIOHDMGLYLNNEATRKLTDDMDILITILYS
GYEVSTKSMAYKVLHDHPKVELEIKREKRAPEPDIDYNDYRSMRTRAVL
LETSLATVNGVLRKTTODMEINGYIIPKMRIVYVTRLEINYPRLPYDPYSNPWR
MDKSLHONSFLEFVGSGTROCPEKELGVAISFELHYFTKYKWEIIGDKLMKPR
VEAPRGLIRISAH"
BASE COUNT 470 a 242 c 268 g 415 t
ORIGIN
alignment_scores:
Quality: 481.00 Length: 469
Ratio: 1.730 Gaps: 13
Percent Similarity: 59.275 Percent Identity: 28.358
alignment_block:
US-09-668-482-2 x SL054770 ..
Align seg 1/1 to: SL054770 from: 1 to: 1395
20 LeuPheLeuAlaValLysLeuTrpGluMetLeuMetIleArgArgVal 36
:::||||| ::|||::| ::|||::| ::|||::| ::|||::|
46 ATCTTTGTACTGCTTATTAAGATGCAATCAAGTATACCAAAA 95
:::|::| ::|||::| ::|||::| ::|||::| ::|||::|
36 lAspProAsnCysArgSerProLeuProGlyThrMetGlyLeuPro 53
:::|::| ::|||::| ::|||::| ::|||::| ::|||::|
96 AAC.....TTGCCCTCGTACTATGAGTGGCCAC 127
53 heIleGlyGluThrLeuGlnLeuIleLeuGlnArgLysPheLeuArg 69
:::|::| ::|||::| ::|||::| ::|||::| ::|||::|
128 TTTTGTGAACACTGAGGTTTCTTAACCTTGCTCAAGTTTCATGAAA 177
70 MetLysArgGlnLysTyrGlyCysIleTyrLysThrHisLeuPheGly 86
:::|::| ::|||::| ::|||::| ::|||::| ::|||::|
178 AACCAAGACCCAGATATGGAGTTTATTAATACACATACTGTTG 227
86 nProThrValArgValMetGlyAlaAspAsnValArgGlnIleLeuLeu 103
|||||::| ::|||::| ::|||::| ::|||::| ::|||::|
228 TCACAAATTTGTTCAATGATTCAGAACTGACAGATATATACTAGTGA 277

```



```

103 1ygluHnHsLysLeuValSerValGlnTrpProAlaSerValArgThrIle 119
    ::||| ||| :::: ||||| |||::: |||
278 ATGAAGCGAAGGACGTGTCGCCAGGATACCCACAGCTCATGATGATGATAT 327
120 LeuGlySerAspThrLeuSerAsnValHisGlyValGlnHisLysAsnLys 136
    |||||::: ::::: |||::: ||| |||||
328 TTAGGAAATGTAATATATGTCAGCTCAATGGTTCCAGCTCAAGTGCAT 377
136 sLysLysAlaIleMetArgAlaPheSer.....ArgAspAlaL 149
    :::: |||::: |||::: ||| |||||
378 GAGGCGTCATTTGTTATCCCTAATTAAGCCCTACAAATGATCAGAACCAAC 427
149 euGlnHisTyrIleProValIleGlnGlnValLysSerAlaIleGln 165
    ::||| |||::: |||::: |||::: |||
428 TT.....TTGCCCTAAATTCATGATGATTATGAGATCCCACTTAAAC 468
166 GluTrpLeuGlnLysAspSerCysValLeuValTyrProGluMetLysLys 182
    ::||| |||::: |||::: |||::: |||
469 AATGGGATAATATAA.....GTTATTGACATTCACAGAGAAACCAATTA 512
182 sLeuMetPheArgIleAlaMetArgIleLeuLeuGlyPheGluProGluG 199
    :::: ||| ::::: |||::: |||::: |||
513 GATGGCATTTCTATCATCGTTCGAAGCAAAATTCGTGATTCGATCTACCT 562
199 InIleLysThrAspGluGlnLeuValGluAlaPheGluGluMetIle 215
    :::: |||::: |||::: ||| ::::: |||
563 CTTTA.....GCTCAGAAATTCATGCTGAAATTTTTCATCATCTAGTG 603
216 LysAsnLeuPheSerLeuProIleAspValProPheSerGlyLeuTyrFar 232
    :::: |||::: |||::: |||::: |||
604 CTAGGCACTCTTCACCTACCTATCATCTCCAAACCAACCATATCATCG 653
232 gGlyLeuArgAlaArg.....AsnPheIleHisSerLysIleG 245
    |||::: |||::: |||::: |||::: |||
654 CGGATTTGAGCAAGCAAAATTCATGCACTTATTACGACACATCATAG 703
245 LuGlu...AsnIleArgLysLysIleGlnAspAsp..... 255
    ||||| :::: |||::: |||::: |||
704 AAGAGAGAGACCTTCAAAGCAAAATTCACATGATATGCTTGTTACCTG 753
256 AspAsnGluAsnGluGlnLysTyrLysAspAlaLeuGlnLeuIleG 272
    |||||::: |||::: |||::: |||
754 ATGAATGAGGAAGCAACACGATTCAA..... 780
272 uAsnSerArgArgSerAspGluProPheSerLeuGlnAlaMetLysGluA 289
    ::||| ||| :::: |||
781 .....TTAACAGATGATGAG.....ATGATTGATT 805
289 laAlaThrGlnLeuLeuPheGlyGlyHisGluThrThrAlaSerThrAla 305
    ::|||::: |||::: |||::: |||::: |||
806 TAATTATTAACCTATTTTGTACTCTGATATGAAGCTTCCACCACTTCT 855
306 ThrSerLeuValMetPheLeuGlnLysAsnThrGluValValGlnLysVa 322
    ||| ::||| :::: |||::: |||::: |||
856 ATGATGGCTGTAATATCTTCATGATCATCCAAAGTCTTGGAAGAACT 905
322 larGluGlu.....ValGlnGluLysValGluMetGlyMetTyrT 336
    ::|||::: |||::: |||::: |||::: |||
906 TAGAAAGAAACACATGGCTATTAGAGAAAAA..... 942
336 hrProGlyLysGlyLeuSerMetGluLeuLeuAspGlnLeuLysTyrThr 352
    ||| :::: |||::: |||::: |||::: |||
943 ..CCTGAGAGATCCTATCGATTCACAAAGATTACAGGCTCAATGCGTTCACA 990
353 gLysCysValIleLysGluThrLeuArgIleAsnProProValProGlyG 369
    ::||| ||| ||||| |||::: |||
991 CGAGCTGTGATTTTAGAGCTCCAGGTTAGCAACAATAGTAATGGGGT 1040
369 yPheArgValAlaLeuLysThrPheGluLeuAsnGlyTyrGlnIleProL 386
    ::||| :::: |||::: |||::: |||::: |||
1041 TTGAGAGAAACAACTCAAGATATGGAATAATGGTATATCATCTCCTTA 1090
386 ySGLYTRPASNVALILETYRSRIIECYSPHTRHISASPYALALASPP 402
    |||||::: |||::: |||::: |||
1091 AAGGATGAGAAATATACGTATATACAAAGGAGTTGCAATTACGATCCACA 1140
403 ValPheProAsnLysGluGluPheGlnProGluArgPheMetSerLysG 419
    ::|||::: |||::: |||::: |||::: |||
1141 CTTATCTGATCCATATTCGTTCAATCCATGAGATGATGATGATGATGAG 1190
419 yLeuGluAspGlySerArgPheAsnTyrIleProPheGlyGlySerA 436
    ::||| :::: |||::: |||::: |||
1191 CCGGAACACCAAAACTCATTT.....TTGGTATTTGGAGGTGTACTTA 1234
436 rGMetCysValGlyLysGluPheAlaLysValLeuLeuLysIlePheLeu 452
    |||::: |||::: |||::: |||::: |||
1235 GACAAATGCTCTGGAAGGAACTTGCTGTAGCAGAAATTTCCACATTTCTT 1284
453 ValGluLeuThrGlnHisCysAsnTrpIleLeuSerAsnGlyProProth 469
    ::||| :::: |||::: |||::: |||
1285 CATTACTTCGTACCAAAATATACAGATGGGAAGAAATAGTGAGATAAAGT 1334
469 rMetLys 471
    |||||
1335 GATGAAA 1341

```


Align seg 1/1 to: AAV09251 from: 1 to: 1850

```

1 MetGlyLeuTyrThrLeuMetValThrPheLeuGlySerThrIleValLeuPro 17
88 AAGGGGCTGTACACCTTATGTGTACCTTCTGTGCACCATCGTGTAC 137
17 OValLeuLeuPheLeuAlaValValLeuTyrPheGluMetLeuMetIle 34
138 CGTTTACTCTTTCTCGCCGGGTGAAGTTGTGGAGATCTTAATGATCC 187
34 rGAGValAspProAsnGlySerSerProLeuProProGlyThrMetGly 50
188 GACGAGTGTCCGAACTGCAAGTCTCTACCGCGCAGTACATGAGGC 237
51 LeuProPheIleGlyGluThrLeuGluMetLeuGluMetLeuArgArgLys 67
238 TTGCGCTTCATTTGGAGAAAGCTCCAGCTGATCCGACAGAGAAAGTT 287
67 eLeuArgMetLysArgGlnLysTyrGlyCysIleTyrLysThrHisLeu 84
288 TCTGCGCATGAAACGCGAGAAATACGGGTGCATCTACAAAGCAGCACTCT 337
84 heGlyAsnProThrValArgValMetGlyAlaAspAsnValArgGlnIle 100
338 TCGGAGACCCGACTGTCAAGGTGATGGAGAGCTGATTAATGAGCAAGAT 387
101 LeuLeuGlyGluHisLysLeuValSerValGlnTrpProAlaSerValAr 117
388 CTGCTGGGCGAACAACAGCTGTGTCTGTTCAGTGGCCAGCATCAGTGG 437
117 gThrIleLeuGlySerAspThrLeuSerAsnValHisGlyValGlnHisL 134
438 AACCATCTGGGCTGTGACACCTCTCCATGTCATGAGATTCACACA 487
134 yAsnLysLysLysAlaIleMetArgAlaPheSerArgAspAlaLeuGlu 150
488 AAAACAAAGAAAAAGCCATATGAGGGGCTCTCCAGATGCTCTGGAG 537
151 HisTyrIleProValIleGlnGlnValLysSerAlaIleGlnGluTr 167
538 CACATCATTTCCCGTATCCAGAGAGAGGTGAAGAGCGCATACAGAGATG 587
167 pLeuGlnLysAspSerCysValLeuValTyrProGluMetLysLysLeu 184
588 GCTGCAGAAAAGACTCCTCGCTGCTGTTATCCAGAAATGAAGAAACA 637
184 ePheArgIleAlaMetArgIleLeuLeuGlyPheGluProGluGlnIle 200
638 TCTTTCGATACCTATGAGAAATCCTGCTGTTTGAACACAGCAATA 687
201 LysThrAspGluGlnGluLeuValGluAlaPheGluGluMetIleLys 217
688 AAGAGCGAGCAGCAAGAACTGCTGGAACCTTTTGAAGAAATGATCAAAA 737
217 rLeuPheSerLeuProIleAspValProPheSerGlyLeuTyrArgGlyL 234
738 CTTGTCTCTCTTCCCAATCGACGTTCTTTCAGTGTCTGTACAGGGGTT 787
234 euArgAlaArgAsnPheIleHisSerLysIleGluGluAsnIleArgLys 250
788 TCGAGGCGACGCAATTTCACTCCAAATGAGAGAAACATCAGAGAG 837
251 LysIleGlnAspAspAsnGluAsnGluGlnLysTyrLysAspAlaLe 267
838 AAAATTCAAGATGACGACATGAAACGAAACGAAATACAAAGAGCGCCT 887
267 uGlnLeuLeuIleGluAsnSerArgArgSerAspGluProPheSerLeu 284
888 TCAGCTGTTGATCGAAGACAGCAAGAAAGTACAGACCTTTAATTTGC 937
284 IAlaMetLysGluAlaAlaThrGluLeuLeuPheGlyGlyHisGluThr 300
938 AGGCGATGAAAGACAGCTACAGAGCTTCTATTGGAGGTCATGAAAC 987

```

```

301 ThrAlaSerThrAlaThrSerLeuValMetPheLeuGlyLeuAsnThrGln 317
988 ACCGCCAGCAGCTCCAACTCATCTTGATGTTTCTGCGTGTGACAGACA 1037
317 uValValGlnLysValArgGluGluValGlnGluLysValGluMetGly 334
1038 AGTGTGCGAAGAGTCAAGAGAGAGGTTCAGGAGAGGTTGAAGTGGCA 1087
334 eTyrThrProGlyLysGlyLeuSerMetGluLeuLeuAspGlnLeuLys 350
1088 TGTATACCTCGTAAGAGGCTTATGATGAGAGCTTTGACACGAGTGAAG 1137
351 TyrThrGlyCysValIleLysGluThrLeuArgIleAsnProProVal 367
1138 TACACTGATGTGTGATTAAAGAGCTTTAGATCAACCTCTCTGTTCC 1187
367 oGlyGlyPheArgValAlaLeuLysThrPheGluLeuAsnGlyTyrGln 384
1188 CGAGGATTCAGAGTCCGACTCAAAACCTTTGATGATGATGATACCAA 1237
1238 TTCCTAAGGATGGAAGCTCATTTACAGCATCTGTGACAGCGAGTGTG 1287
401 AlaAspValPheProAsnLysGluGluPheGlnProGluLeuPheMet 417
1288 GCGGACCTCTTCCAAACAAAGAGAGTTCACGCCGAGAGATTCATGAG 1337
417 rLysGlyLeuGluAspGlySerArgPheAsnTyrIleProPheGlyGly 434
1338 CAAAGCTGTGAGAGCGGGTCCAGGTTTAACTACATCCCTTGGAGAG 1387
434 LysSerArgMetCysValGlyLysGluPheAlaLysValLeuLeuLys 450
1388 GATCCAGATGTGTGTGGCAAGAGTTCGCCAAAGTGTACCAAGATC 1437
451 PheLeuValGluLeuThrGlnHisCysAsnTrpIleLeuSerAsnGly 467
1438 TTTTACTTGAATTAAAGCAGCATTCGAATTCGATTCCTCAAAACGAGC 1487
467 oProThrMetLysThrGlyProThrIleTyrProValAspAsnLeuPro 484
1488 CCGGACATGAAACAGGCCGACTATTTTACCAAGTGGACAACTCTCCT 1537
484 hLysPheThrSerTyrValArgAsn 492
1538 CCAATTCACACTGATTATGTCAGAAAT 1563

```

seq_name: /STDS1/gcgcdata/geneseq/geneseqn/NA1998.DAT:AAV12203

seq_documentation_block:

ID AAV12203 standard; cDNA; 1850 BP.

XX AAV12203;

XX 22-JUN-1998 (first entry)

XX DE zebrafish retinoid metabolising protein zp450RAI cDNA.

XX KW retinoid metabolising protein; zp450RAI; retinoid oxidase;

XX KW retinoid acid; zebrafish; inhibitor; antisense; cancer;

XX KW actinic keratosis; oral leukoplakia; head tumour; neck tumour;

XX KW non-small cell lung carcinoma; basal cell carcinoma;

XX KW acute promyelocytic leukaemia; skin cancer; acne; psoriasis;

XX KW Ichthyosis; therapy; diagnosis; screening; ss.

XX Danto rerio.

XX OS

XX FH key Location/Qualifiers

XX CDS 88..1566

XX FT /tag= a

XX

PN WO9749815-A1.
 XX 31-DEC-1997.
 PD 23-JUN-1997; 97WO-CA00440.
 XX 01-OCT-1996; 96US-0724466.
 PR 21-JUN-1996; 96US-0667546.
 XX (TOOH) UNIV QUEENS KINGSTON.
 PA Beckett BR, Jones G, Petkovich PM, White JA;
 PI WPI: 1998-07178/07.
 DR P-PSDB; AAM12203.
 XX Retinoid metabolising protein - useful to develop products to treat,
 PT e.g. cancer, actinic keratosis, oral leukoplakia, acne, psoriasis or
 PT ichthyosis
 PS Claim 15; Page 52-54; 110pp; English.
 XX This cDNA clone codes for a novel zebrafish retinoid metabolising
 CC protein (see AAM4153) designated zp450RA1. The encoded protein is a
 CC retinoid oxidase that has the ability to hydroxylate retinoic acid
 CC at the 4 position of the beta-ionone ring and is inducible in
 CC epithelial cells exposed to retinoic acid. To isolate the clone,
 CC transcripts present in fin tissue regenerating in the presence or
 CC absence of retinoic acid were compared using a differential display
 CC PCR technique, and an isolated clone (see AAV12213) was used to screen
 CC a D. rerio 6-18 hr embryo cDNA library. Human (see AAV12204) and
 CC mouse (see AAV12205) P450RA1 clones have also been identified. The
 CC isolated clones can be used in the recombinant production of
 CC P450RA1 proteins. Antisense nucleic acids can be used in a claimed
 CC method for inhibiting retinoic acid hydroxylation for the treatment
 CC of cancer, actinic keratosis, oral leukoplakia, secondary tumours of
 CC the head and/or neck, non-small cell lung carcinomas, basal cell
 CC carcinomas, acute promyelocytic leukaemia, skin cancer, and
 CC premalignancy associated with actinic keratosis, acne, psoriasis
 CC and/or ichthyosis. Promoter sequences (see AAV12206-08) are also
 CC claimed.
 XX
 XX Sequence 1850 BP; 519 A; 391 C; 454 G; 486 T; 0 other:
 SQ
 alignment_scores:
 Quality: 2563.00 Length: 492
 Ratio: 5.209 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000
 alignment_block:
 US-09-668-482-2 x AAV12203 ..
 Align seg 1/1 to: AAV12203 from: 1 to: 1850
 1 MetGlyLeuTyrThrLeuMetValThrPheLeuGlySThrIleValLeuP 17
 88 AAGGGGCTGTACACCTTATGGTCACTTCTCTGACCATGTGCTACCC 137
 17 oValLeuLeuPheLeuAlaIleValIysLeuTyrPgluMetLeuMetIleA 34
 138 CGTTTACTCTTCTCGCGCGGTGAAGTGTGGAGATGTTAATGATGCC 187
 34 rGArgValAspProAsnGlySerProLeuProProGlyThyMetGly 50
 188 GACGAGTCATCCGAACTGCAGAACTGCTCTACCCGACGTCATGCGC 237
 51 LeuProPheIleGlyLthrLeuGlnLeuIleLeuGlnArgIysPhe 67
 238 TTGCGCTTATTGGAGAAAGCTCCAGCTGATCTCCAGAGAAAGAGTT 287
 67 eLeuArgMetIysArgGlnIysTyrGlyCysIleTyrIysThrHisLeuP 84

288 TCTGGCATGAACGGCAGAAATACGGGTGCATCTACAAAGCAGCTCT 337
 84 heGlyAsnProThrValArgValMetGlyAlaAspAsnValArgGlnIle 100
 338 TCGGGAACCCGACTGTCAAGGTGATGGAGGTGATGAATGTGAGCAGATT 387
 101 LeuLeuGlyGluHisIysLeuValSerValGlnTyrProAlaSerValAr 117
 388 CTGCTGGGCGAACAAGCTGGTGTCTTTCAGTGGCCAGCTCAGTGGAG 437
 117 gThrIleLeuGlySerAspThrLeuSerAsnValHisGlyValGlnHisL 134
 438 AACCATCTGGGCTGTGACACCTCTCCAAATGTCATGAGTCAACACA 487
 134 yAsnIysIysIysAlaIleMetArgAlaPheSerArgAspAlaLeuGlu 150
 488 AAACAAAGAAAAAGGCATTATGAGGCGCTTCCTCGAATCTCTGGAG 537
 151 HsTyrIleProValIleGlnGlnIleValIysSerAlaIleGlnGluTr 167
 538 CACTACATTCCCTGATCCAGCAGAGGTGAAGGCGCATACAGGAATG 587
 167 PleuGlnIysAspSerCysValLeuValTyrProGluMetIysIysLeuM 184
 588 GCTGCAAAAAAGACTCCTCGCTGCTGTTATCCAGAAATGAAGAACTCA 637
 184 eThearGylleAlaMetArgIleLeuLeuGlyPheGluProGluGlnIle 200
 638 TGTTTGGATGATGAGAAATCTCTGTTGTTGAACAGCAGCAAAATA 687
 201 LysThrAspGluGlnIleLeuValGluAlaPheGluGluMetIleIysAs 217
 688 AAACGAGCAGCAAGAACGTCGTGAAGCTTTGAGGAATGATCAAAAA 737
 217 nLeuPheSerLeuProIleAspValProPheSerGlyLeuTyrArgGlyL 234
 738 CTGTGCTCTCTCCCAATGACGCTTCCTTCACTGAGTGTGTACAGGGGTT 787
 234 euArgAlaArgAsnPhelIleHisSerIysIleGlnGluAsnIleArgIys 250
 788 TGAAGGCGACCAATTTCATTCACCTCCAAATTAGCAAAACATCAGGAAG 837
 251 LysIleGlnAspAspAsnGluAsnGluGlnIysTyrIysAspAlaLe 267
 838 AAATTCACATGACGACCAATGAAGACGAAACAGAAATACAAAGCCCT 887
 267 uGlnLeuLeuIleGluAsnSerArgArgSerAspGluProPheSerLeu 284
 888 TCAGCTGTGATCGAAGAACAGCAAGAAAGTGAACCTTTAGTTTGC 937
 284 lnaIleMetIysGluAlaAlaThrGluLeuLeuPheGlyGlyHisGluThr 300
 938 AGCGGATGAAGAAAGACGCTACAGAGCTTCTATTGGAGTCAATGAAC 987
 301 ThrAlaSerThrAlaThrSerLeuValMetPheLeuGlyLeuAsnThrG 317
 988 ACCGCCAGCACTGCACCTCATCTGTCTCTGCGTCTGAACACAGA 1037
 317 uValValGlnIysValArgGluGluValGlnIleValGlnMetGlyM 334
 1038 AGTGTGTGAGAAAGGTTCAGAGAGAGGTTTCAGGAGAAAGTTGAATGGGCA 1087
 334 eTyrThrProGlyIysGlyLeuSerMetGluLeuLeuAspGlnIleIys 350
 1088 TGTATACACCTCGAAAGGCTTGAATGAGAGCTGTTGGACAGAGCTGAAG 1137
 351 TyrThrGlyCysValIleIysGluThrLeuArgIleAsnProProValPr 367
 1138 TACACTGGATGTGATTAAGAGACTCTTAGAATCAACCTCTGTGTC 1187
 367 oGlyGlyPheArgValAlaLeuIysThrPheGluLeuAsnGlyTyrGlnI 384
 1188 CGAGGATTCAGAGTGCACCTCAAAACCTTTGAATGAAATGTTACCAA 1237

```

384 leProlySGlyTyrPasnValIleTyrSerIleCysAspThrHisAspVal 400
|||||
1238 TTCCTAAGAGATGGAACCTCATTTTACACATCTGTGACACGACGATGTG 1287
401 AlaAspValPheProAsnLysGluGluPheGlnProGluTyrPheMetSe 417
|||||
1288 GCCGACGCTTCCCAACAAAGAGAGTTCCAGCCGAGAGATTCATCAG 1337
417 fLySGlyLeuGluAspGlySerArgPheAsnTyrIleProPheGlyGly 434
|||||
1338 CAAAGGCTGTGAGGACGCGTCCAGGTTTAACATCACTCCCTCGAGAGAG 1387
434 lYSerArgMetCysValAlGlyLysGluPheAlaLysValLeuLeuLysIle 450
|||||
1388 GATCCAGAGATGTGTGGGCAAAAGATTCGCCAAAGTTCCTCAAGATC 1437
451 PheLeuValGluLeuThrGlnHisCysAsnTrpIleLeuSerAsnGlyPr 467
|||||
1438 TTTTAAAGTTGAGTTACGACGACATTCGCAATTCGATTCCTCAAGGACC 1487
467 oProThrMetLysThrGlyProThrIleTyrProValAspAsnLeuProT 484
|||||
1488 CCCGACATATAAAACAGGCCCGAGCTATTACCCAGTGAGACATCTCCCTA 1537
484 hriLysPheThrSerTyrValArgAsn 492
|||||
1538 CCAAAATTCACCTAGTTATGTCAGAAAT 1563

seq_name: /std1/gcgsdata/geneseq/geneseqn/NA1998.DAT:AAV09247
seq_documentation_block:
ID AAV09247 standard; cDNA: 1494 BP.
XX AAV09247;
AC
XX
XX
XX 07-JUL-1998 (first entry)
XX
DE Human cytochrome P450RA1 cDNA sequence.
XX
XX Retinoid regulated gene; cytochrome P450 gene; enzyme:
XX oxidative metabolism; P450RA1; retinoic acid; RA; promoter; ss.
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 1..1494
XX FT /*tag= a
XX FT /product= "Cytochrome P450RA1"
XX
XX WO9749832-A2.
XX
XX 31-DEC-1997.
XX
XX 23-JUN-1997; 97WO-CA00488.
XX
XX 01-OCT-1996; 96US-0724466.
XX 21-JUN-1996; 96US-0667346.
XX (TOOH ) UNIV QUEBENS KINGSTON.
XX
XX Petkovich PM;
XX
XX WPI: 1998-077193/07.
XX P-PSDB: AAW37734.
XX
XX Identifying DNA encoding inducible or suppressible cytochrome P450 -
XX by screening for drugs which reduce the catabolism of retinoic acid,
XX useful in cancer chemotherapy and the treatment of acne and
XX psoriasis
XX
XX Example 4; Pages 56-58; 113pp; English.
XX

```

```

CC This nucleotide sequence encodes the human cytochrome P450RA1. Its
CC expression is dependent on the presence of retinoic acid (RA). The
CC retinoid-regulated genes such as the inducible cytochrome P450RA1
CC gene specifically metabolises a derivative of the RA. The cytochrome
CC P450 gene in general produces enzymes involved in the oxidative
CC metabolism of endogenous and exogenous compounds. The cytochrome
CC P450 nucleotide sequence can be used to induce or suppress the
CC expression of its protein. P450RA1 is highly induced by RA in cell
CC lines and tissues. This allows for development of a drug screen
CC using promoters and nucleotide sequences to identify drugs which are
CC useful for reducing the catabolism of RA.
XX
XX Sequence 1494 BP; 332 A; 414 C; 430 G; 318 T; 0 other;

alignment_scores:
Quality: 1738.00 Length: 497
Ratio: 4.080 Gaps: 5
Percent Similarity: 85.714 Percent Identity: 68.008

alignment_block:
US-09-668-482-2 x AAV09247 ..
Align seg 1/1 to: AAV09247 from: 1 to: 1494

1 MetGlyLeuTyrThrLeuMetValThrPheLeuCysThrIleValLeuPr 17
|||||
1 ATGGGGCTCCCGCGCTGCGCCAGTGCGCTGCACCTTCGCTGCTGCC 50
17 OValLeuLeuPheLeuAlaAlaValLysLeuTrpGluMetLeuMetIleA 34
|||||
51 GCTGCTCTCTCTCTGCTGCTGATCAAGCTTGCGACTGACTGCTGA 100
|||||
34 rGaTgValAspProAsnCysArgSerProLeuProProGlyThrMetGly 50
|||||
101 GCGGCCCGACCGCAGCTGTGCTCCCATTCGCCCCCGGAGCTATGGGC 150
51 LeuProPheIleGlyLysIleThrLeuGlnLeuIleLeuGlnArgArgLysph 67
|||||
151 TTCCTCTTCTTTGGGAACCTTGCAATGATGTACTGAGCGGAGAAAGTT 200
67 eLeuArgMetLysArgGlnLysTyrGlyCysIleTyrLysThrHisLeuP 84
|||||
201 CCTGACATGAAGCGCAGAAATACGGCTTCATCTACAACAGCATGTGT 250
84 heGlyAsnProThrValArgValMetGlyAlaAspAsnValArgGlnIle 100
|||||
251 TCGGGGCGCCACCGTACGGGTATGGGCGGAGCAATGTGCGCGCATC 300
101 LeuLeuGlyGlnHisLysLeuValSerValGlnTrpProAlaSerValAr 117
|||||
301 TTGCTCGGAAGACGACGCGTGTGCTGCTCACTGGCGACGTCGGTGGC 350
117 gThrIleLeuGlySerAspThrLeuSerAsnValHisGlyValGlnHisL 134
|||||
351 CACCATTTGGGATCTGGCTGCTCTTAACCTGCACGACTCTCCACACA 400
134 yAsnLysLysLysAlaIleMetArgAlaPheSerArgAspAlaLeuGln 150
|||||
401 AGCAGCGCAAGAAAGTGATTAATCGGCGCTTCAGCCGAGGACATCGAA 450
151 HisTyrIleProValIleGlnGlnGluValLysSerAlaIleGlnLys 167
|||||
451 TGCTACGTGCGGCTGATCAACGAGAAAGTGGGAGCAGCGCTGAGAGAG 500
167 pleuGlnLysAspSerCys.....ValLeuValTyrProGluM 180
|||||
501 GCTG.....AGCTGGCGAGCGCGCCCTCCGTGCTACCCCGAGG 541
180 eTylLysLeuMetPheArgIleAlaMetArgIleLeuLeuGlyPheGlu 196
|||||
542 TGAAGGCTCATGTTCCGAATCGCATCGCATCTACTGGGCTGCGAA 591

```



```

1 MetGlyLeuTyrThrLeuMetValThrPheLeuGlyThrIleValLeuPr 17
1 ATGGGGCTCCCGGCGCTGCTGCGCAGTCCGCTGCGACCTTCTGCTGCC 50
17 ovalLeuPheLeuAlaValLysLeuTrpGluMetLeuMetIleA 34
1 GCTGCTGCTTCTCCGCTGCGATCAACCTGCGGACCTGCTACTGCGGA 100
34 rgaArgValAspProAsnGlySerProLeuProGluTyrMetGly 50
101 GCGGCGCGGACCGACGATGTCCTCCATTCGCCGCCCGGACCTATGGCC 150
51 LeuProPheIleGlyGluThrLeuGluLeuIleLeuGlnArgArgLysP 67
151 TTCCCTCTTCTTGGGGAACCTTGACATGATGCTACGCGAGGAAGTT 200
67 eLeuArgMetLysArgGlnLysTyrGlyCysIleTyrLysThrHisLeu 84
201 CCGACGATGAAGCGGAAATACGCTTCTATCAAGACGACTCTGT 250
84 heGlyAsnProThrValArgValMetGlyAlaAspAsnValArgGlnI 100
251 TCGGCGGCGCCACCGTACGGGATGGCGGCGGACATGCGGCGCATC 300
101 LeuLeuGlyGluHisLysLeuValSerValGlnTrpProAlaSerVal 117
301 TTGCTCGAGAGACGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 350
117 gThrIleLeuGlySerAspThrLeuSerAsnValHisGlyValGlnHis 134
351 CACCATTTCTGGGATCTGCTGCTGCTTCTACCTGACGACCTCCGCGCA 400
134 yAsnLysLysLysAlaIleMetArgAlaPheSerArgAspAlaLeuG 150
401 AGCAGCGGAGAGAGGTATATGCGGCGCTTACGCGGAGGACACTCGAA 450
151 HisTyrIleProValIleGlnGlnValLysSerAlaIleGlnGluTr 167
451 TGCTACGCTGCGCTGATCAACGAGAACTGGCGACGACCTGGAGCATG 500
167 pLeuGlnLysAspSerCys.....ValLeuValTyrProGluM 180
501 GCTG.....AGCTGCGGCGAGCGGCGCTCTCTGCTTACCCGCGAG 541
180 eLysLysLeuMetPheArgIleAlaMetArgIleLeuLeuGlyPheGlu 196
542 TGAAGCGCTCATGTTCCGAATCGCATGCTCTACTGCGCTGCGGAA 591
197 ProGluGlnIleLysThrAsp.....GluGlnGluLeuValGluAl 210
592 CCC.....CAACTGCGGCGAGCGGAGCTCCGAGCAGCAGCTTGAGAGC 638
210 aPheGluGluMetIleLysAsnLeuPheSerLeuProIleAspValPro 227
639 CTTGCGAGAAATGACCGCAATCTTCTCTGCTGCGCAATCGAGCGGCC 688
227 heSerGlyLeuTyrArgGlyLeuArgAlaArgAsnPheIleHisSerLys 243
689 TCAGCGGCTGCTACCGGCGCATGAGCGCGGAACTCATTCACGCGCG 738
244 IleGluGluAsnIleArgLysLysIleGlnAsp...AspAsnGluAs 259
739 ATCGAGCGCAACATTCGCGCCAAAGATCTGCGGCGCTGCGGCGATCCGAGGC 788
259 nGluGlnLysTyrLysAspAlaLeuGlnLeuLeuIleGluLysSerArg 276
789 GGGCGAGGCGCTCGCAAGAGCGCGCTGCTGCTGCTGCTGCTGCTG 838
276 rGSerAspGluProPheSerLeuGlnAlaMetLysGluAlaAlaThrGlu 292
839 AGAGGAGAGAGCGCTGCGCATGCGACATGAAGCAATCTTCAACCGAA 888

```

```

293 LeuLeuPheGlyGlyHisGluThrThrAlaSerThrAlaThrSerLeuVa 309
889 CTCCTCTTTGAGAGACACGAAACACGCGCAGTGCAGCAGCAATCTGTAT 938
309 lMetPheLeuGlyLeuAsnThrGluValAlaGlnLysValArgGluGly 326
939 CACTTCTGCGGCGCTTACCCACATGTTCTCCAGAAAGTGGAGAGAGC 988
326 aGlnGlnLysValGluMetGlyMetTyrThrProGlyLysGlySer 342
989 TGAAGAGTAGAGTTTACTTTCAGAGCAATCAAGCAACAAGTTGGAGC 1038
343 MetGluLeuLeuAspGlnLeuLysTyrThrGlyCysValIleLysGlu 359
1039 ATGGAATTTTGGACAACTTAATATCATCGGCTGTGTTAATGAAGAGC 1088
359 rLeuArgGlyLeuAsnProProValProGlyGlyPheArgValAlaLeuLys 376
1089 CTTTGCACCTAATCCCGCAGTTCAGAGAGGTTTCGGGTTGCTGTGAA 1138
376 hrPheGluLeuAsnGlyTyrGlnIleProLysGlyTyrPasnValIle 392
1139 CTTTGAATTAATGAATACCAATTCAGAGGCTGAGATGTTATCTAC 1188
393 SerLleCysAspThrHisAspValAlaAspValPheProAsnLysGlu 409
1189 AGTATCTGTATCTCATGATGTGCGAGAGATCTTACCAACAGAGAA 1238
409 upheGlnProGluArgPheMetSerLysGlyLeuGlnAspGlySerArg 426
1239 ATTAAATCCGACGAGATTCAGTCTCTCTCAACCGAGAGATGCTTCAG 1288
426 heAsnTyrIleProPheGlyGlyLysSerArgMetCysValGlyLysGlu 442
1289 TCAGCTCATTTCAATTTGGAGAGGCTTACAGAGCTGTGAGCAAGAA 1338
443 PheAlaLysValLeuLeuLysIlePheLeuValGluLeuThrGlnHis 459
1339 TTTCGAAATAATCTTCAATAATATTACAGTGAGCTGGCCAGCATTC 1388
459 sAsnTrpIleLeuSerAsnGlyProProThrMetLysThrGlyProThr 476
1389 TGACTGCGACCTCTTAATGAGCTCTTACAAATGAATAACCAAGTCCAC 1438
476 lTyrProValAspAsnLeuProThrLysPheThrSerTyr 489
1439 TGTATCTGTGAGCAATCTCTCGCAAGATTCACCATTTTC 1479
seq_name: /STD1/gcgdata/geneseq/geneseqn/MA1998.DAT:AAV09252
seq_documentation_block:
ID AAV09252 standard; cDNA; 1725 BP.
XX
AC AAV09252;
XX
DT 07-JUL-1998 (first entry)
XX
DE Cytochrome P450RA1 isoform cDNA sequence.
XX
KW Retinoid regulated gene; cytochrome P450 gene; enzyme;
oxidative metabolism; P450RA1; retinoic acid; RA; promoter; ss;
isoform.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT CDS 24..1518
FT /tag= a
FT /product= "Cytochrome P450RA1 isoform"
XX
PD MO9749832-A2.
XX
XX 31-DEC-1997.

```


XX 23-JUN-1997: 97WO-CA00488.
 PF 01-OCT-1996: 96US-0724466.
 XX 21-JUN-1996: 96US-0667546.
 PR (TOOH) UNIV QUEENS KINGSTON.
 PA
 PI Petkovich PW;
 XX WPI: 1998-077193/07.
 XX P-PSDB; AAW37735.
 DR
 XX Identifying DNA encoding inducible or suppressible cytochrome P450 -
 PT by screening for drugs which reduce the catabolism of retinoic acid,
 PT useful in cancer chemotherapy and the treatment of acne and
 PT psoriasis
 PS Disclosure; Pages 59D-59F; 113pp; English.
 CC This nucleotide sequence encoding an isoform of cytochrome P450RA1.
 CC Its expression is dependent on the presence of retinoic acid (RA).
 CC The retinoid-regulated genes such as the inducible cytochrome P450RA1
 CC gene specifically metabolises a derivative of the RA. The cytochrome
 CC P450 gene in general produces enzymes involved in the oxidative
 CC metabolism of endogenous and exogenous compounds. The cytochrome P450
 CC nucleotide sequence can be used to induce or suppress the expression
 CC of its protein. P450RA1 is highly induced by RA in cell lines and
 CC tissues. This allows for development of a drug screen using promoters
 CC and nucleotide sequences to identify drugs which are useful for
 CC reducing the catabolism of RA.
 XX
 SQ Sequence 1725 BP; 408 A; 425 C; 475 G; 417 T; 0 other;
 alignment_scores:
 Quality: 1734.00 Length: 496
 Ratio: 4.070 Gaps: 4
 Percent Similarity: 85.887 Percent Identity: 67.742
 alignment_block:
 US-09-668-482-2 x AAV09252 ..
 Align seg 1/1 to: AAV09252 from: 1 to: 1725
 1 MetGlyLeuTyrThrLeuMetValThrPheLeuCystrHisLeuValLeu 17
 25 ATGGGGCTCCGGCGCTGCTGCGCAAGTGCGCTCGACCTGCTGCTGCGC 74
 17 oValLeuLeuPheLeuAlaValAlaLysLeuTrpGluMetLeuMetIle 34
 75 GCTGCTGCTCTTCTGCGCGCTCAAGCTCGGAGCACTGTACTGTGTGA 124
 34 rGArGValAspProAsnGlySerProLeuProGlyLysMetGly 50
 125 GCAAGCGGCACTGCACTGCGCTTCTGCGCGCGGTAACATGAGC 174
 51 LeuProPheIleGlyLysLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 67
 175 TTCCCATTTCTTTGGGGAACATTCAGATGATGCTTCACGAGAGAAATT 224
 67 GluLeuArgMetLysArgGlnLysTyrGlySerIleTyrLysThrHisLeu 84
 225 TCTGCGAGATGAACCGCAAAATACGGCTTCATTCACAGACGCACTGT 274
 84 heGlyAsnProThrValArgValMetGlyAlaAspAsnValArgGlnIle 100
 275 TTGGCGCGCCACGATGCGGATGATGGCGCGGATATGTCGCGCGCAATC 324
 101 LeuLeuGlyLysHisLysLeuValSerValGlnTrpProLysSerValAr 117
 325 TTGCTGGAGACGACCGGTTGCTGCTGCACTGCGCGCGTGGTGGC 374

117 gThrIleLeuGlySerAspThrLeuSerAsnValHisGlyValGlnHisL 134
 375 CACCATCTGGGCGCTGCTGCTTCCAACTGCAAGATTCCTGCGCA 424
 134 yAsnLysLysAlaIleLeuArgAlaPheSerArgAspAlaLeuGln 150
 425 AGCAGGAAAGAGTATATGACAGCCTTCAGCGGAGGACACTCGAG 474
 151 HisTyrIleProValIleGlnGlnValLysSerAlaIleGlnLys 167
 475 TGCTACGTGCTGCTGATCCGTGAGGAGTACGATGTCGAGGAGT 524
 167 PleuGlnLysAspSerCys.....ValLeuValTyrProGluM 180
 525 GCTA.....AGCTGCGGAGCGCGGCTCTGCTGCTACCCGAG 565
 180 eLysLysLeuMetPheArgIleAlaMetArgIleLeuLeuGlyPheGln 196
 566 TGAAGCGCTCATGTTCCGATCGCATCGCATCTGCTGGCGTCCGAG 615
 197 ProGlnGlnIle.....LysThrAspGlnGlnLeuValGlnLap 211
 616 CCGGCTCAGCGGCGCGGAGAGAGCAACAGCTGCTGAGAGCTT 665
 211 eGluGluMetIleLysAsnLeuPheSerLeuProIleAspValProPhe 228
 666 CGAGAGATGACCCGCAATCTTCTCTCTCCATGACGTGCCCTTTA 715
 228 eArgLysLeuTyrArgGlyLeuArgAlaArgAsnPheIleHisSerLys 244
 716 GCGGCTGTACCGGCGGCGTGAAGCGCGCAACTTATACAGCGCGCATC 765
 245 GluLysAsnIleArgLysLysIleGlnAsp...AspAspAsnGln 260
 766 GAGGAGAACATTCGCGCAAGATCCGCGCTTACGCTACAGAGCGGA 815
 260 uGlnLysTyrLysAspAlaLeuGlnLeuLeuIleGlnLysSerArgArg 277
 816 TGGGGGTTCGAAAGACGCGCTCGACGCTCTGATGACACTCGGGGAGA 865
 277 eAspGluProPheSerLeuGlnAlaMetLysGlnAlaIleThrGlnLeu 293
 866 GGGAGAGAGCTGATATGACGACACTAAACATGCTCAACAGACCTC 915
 294 LeuPheGlyLysHisGlnThrThrAlaSerThrAlaThrSerLeuVal 310
 916 CTCTTGGTGTCAAGAACTACAGCCAGTGTCTGCGACATCAGATCAC 965
 310 PheLeuGlnLysAsnThrGlnValAlaGlnLysValArgGlnLysVal 327
 1016 AGAGCAAGCGCTTACTTTCAGAGCAATCAAGCAACAACTTGAAGATG 1065
 344 GluLeuLeuAspGlnLeuLysTyrThrGlyCysValIleLysGlnThr 360
 1066 GAAACTTGGAAACGCTTAATATACATTGGGTGTGTGATTAAGAGACCT 1115
 360 uArgIleAsnProProValProGlyGlyPheArgValAlaLeuLysThr 377
 1116 GCGATTGAATCTCGGTTCCAGAGGAGGTTTCGGGTCTCTGAAGACTT 1165
 377 heGlnLeuAsnGlyTyrGlnIleProLysGlyTyrPheValIleTyrSer 393
 1166 TTGACCTGAATGATACAGATCCCAAGGCGTGAAGTATTATTACGT 1215
 394 IleCysAspThrHisAspValAlaAspValPheProAsnLysGlnLys 410
 1216 ATCTGTGACACCCAGATGTGCAATATCTTCACTAACAGAGGAATTT 1265
 410 eGlnProGluArgPheMetSerLysGlyLeuGlnLysPheArgPhe 427


```

616 CCGGTCACCGGGCGGGAGAGACAGCAAGCTCGTGAGGCTTT 665
211 egluGlMetIleuYsasnleupheserleuprolleaspyalProPhes 228
||||| ..... |||||||
666 CGAGAGATACCCCGCATCTCTCTCTCCATTGACGTCCCTTTA 715
228 erGlyLeuYrArgGlyLeuArgAlaArgAsnphelIehISerLysIle 244
||||| ..... |||||||
716 GCGGCTGTACCGGGCGGTGAGCGGGAACCTTATACAGCGCGCATC 765
245 GluGluasnIleargLysIleGlnAsp...AspAspAsnGluasnG 260
||||| ..... |||||||
766 GAGAGAGACATTCGCCCAAGATCCGCCCTTCAGCTTCAGAGCCGGA 815
260 uGlnLysTyrIleAspAlaLeuGlnleuIleGluAsnSerArgArgS 277
||||| ..... |||||||
816 TGGGGGTTCGAGAGACCGCGCTGCGAGCTCTGATGACACCTCGGGAGA 865
277 erAspGluProPheserleuGlnAlaMetLysGluAlaIleThGluLeu 293
||||| ..... |||||||
866 GGGAGAGAGGCTGATATGAGGCACTAAACATATGCTCAACAGAGCTC 915
294 LeupheGlyGlyIleGluThrAlaSerThrAlaThrSerleuValMe 310
||||| ..... |||||||
916 CTCTTGGTGTGATGAAACTACAGCCAGTCTCGACATCATGATCAGAC 965
310 tPheLeuGlyLeuAsnThrGluValAlaGlnLysValArgGluValG 327
||||| ..... |||||||
966 TTACTAGTACATCCACCATGTCCTCCAGAAAGTTCGAGAGAGATAA 1015
327 InGluLysValGluMetGlyMetYrThrProGlyLysGlyLeuSerMet 343
||||| ..... |||||||
1016 AGAGCAAGGCTTACTTTCAGAGACATCAAGACACATGTTAGACATG 1065
344 GluLeuLeuAspGlnLeuLysTyrThrGlyCysValIleLysGluThrLe 360
||||| ..... |||||||
1066 GAAACTTTGGAAAGCTTAATATCTTGGGTGTGCATTAAAGAGACCTT 1115
360 uArgIleAsnProProValProGlyGlyPheArgValAlaLeuLysThrP 377
||||| ..... |||||||
1116 GCGATTGAATCTCCGCTCCAGAGGGGTTCCGGTCTCTCGAAGACTT 1165
377 heGluLeuAsnGlyTyrGlnIleProLysGlyTyrAsnValIleYrSer 393
||||| ..... |||||||
1166 TTGACCTGATGATACAGATCCCAAGGGCTGGAATGTTATTACAGT 1215
394 IleCysAspThrHisAspValAlaAspValPheProAsnLysGluGluPh 410
||||| ..... |||||||
1216 ATCTGTGACACCCAGATGTGCAGATATCTTCACTAACAAAGAGAAATT 1265
410 eGlnProGluArgPheMetSerLysGlyLeuGluAspGlySerArgPheA 427
||||| ..... |||||||
1266 TTAATCCGACCCCTTATATGTCCTCATCCAGAGATGCTCCCGGTTCA 1315
427 snTyrIleProPhesGlyGlySerArgMetCysValGlyLysGluPhe 443
||||| ..... |||||||
1316 GCTTCATTCATTTGGAGAGCGCTTCGAGCTGTGTAGGCAAAAGATT 1365
444 AlaLysValLeuLeuLysIlePheLeuValGluLeuThrGlnHisCysAs 460
||||| ..... |||||||
1366 GCAAAATATCTTCTTAAGATATTACAGTGAAGCTGTAGGCACTGTGA 1415
460 nTyrIleLeuSerAsnGlyProProThrMetLysThrGlyProThrIleT 477
||||| ..... |||||||
1416 TTGGCAGCTTCTTAATGAGACTCTCTCAATGAAGACAGCCCACTGTGT 1465
477 YrProValAspAsnLeuProThrLysPheThrSerTyr 489
||||| ..... |||||||
1466 ACCCTGTGGACATCTCCCTGCAAGATTCACCTACTTC 1503
seq_name: /SIDSI/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA60752
seq_documentation_block:

```

```

ID AAA60752 standard; cDNA; 2124 BP.
XX
AC AAA60752;
XX
PT 31-OCT-2000 (first entry)
XX
DE Human PSEC64 encoding nucleotide sequence SEQ ID NO:1.
XX
KW Human; PSEC64; neuron growth; nerve disease; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 46..696
FT     /tag= a
FT     /product= "PSEC64"
FT     /note= "protein related to neuron growth"
XX
XX
XX JP2000152790-A.
XX
XX 06-JUN-2000.
XX
XX 19-NOV-1998; 98JP-0329989.
XX
XX 19-NOV-1998; 98JP-0329989.
XX
XX (HER1-) HERIKKUSU KENKIYUSHO KK.
XX
XX WPI; 2000-468126/41.
XX
XX P-PSDB; AAB12489.
XX
XX A protein related to the growth of neuron and a gene encoding said
XX protein -
XX
XX Claim 2; Page 8-10; 13pp; Japanese.
XX
XX The present sequence encodes a human protein, designated PSEC64, which
XX is related to neuron growth. The PSEC64 protein and its gene can be used
XX for the development of a preventive agent for use in the treatment of
XX diseases in which nerves are involved.
XX
XX Sequence 2124 BP; 509 A; 510 C; 547 G; 558 T; 0 other;

alignment_scores:
Quality: 1725.00 Length: 497
Ratio: 4.040 Gaps: 5
Percent Similarity: 85.915 Percent Identity: 68.612

alignment_block:
us-09-668-482-2 x AAA60752 ..
Align seg 1/1 to: AAA60752 from: 1 to: 2124

1 MetGlyLeuTyrThrLeuMetValThrPheLeuCysThrIleValLeuPr 17
||||| ..... |||||||
46 ATGGGGCTCCCGGCTGCTGCGCCAGTGCCTGCTGCACTTCGTCGTGCC 95
17 ovalLeuLeuPheLeuAlaIleValLysLeuThrGluLeuMetIleA 34
||||| ..... |||||||
96 GCTGTGCTCTTCTGCTGCTGCGATTAAGCTCTGGAGCTGTACTGCGGCA 145
146 GCGGGCGGAGCGGAGTTGTGCTGCCATTCGCCCGGAGACTATGGGC 195
34 rGArgValAspProAsnCysArgSerProLeuProProGlyThrMetGly 50
||||| ..... |||||||
146 GCGGGCGGAGCGGAGTTGTGCTGCCATTCGCCCGGAGACTATGGGC 195
51 LeuProPheIleGlyGluThrLeuLeuLeuLeuLeuGlnArgArgLysPh 67
||||| ..... |||||||
196 TTCCCTTCTTGTGGGAAACCTTGCAAGATGACTGCAAGCGAGAGACTT 245
67 eLeuArgMetLysArgGlnLysTyrGlyCysIleTyrIleYrThrHisLeuP 84
||||| ..... |||||||
246 CTTGCAGATGAAGCGCAGAAATACGGCTTCACTACAAAGACGATCTGT 295

```

```

84 heGlyAsnProThrValArgValMetGlyAlaAspAsnValArgGlnIle 100
|||||
296 TCGGGCGCCACCGTACGGGTGATGGCGCGGACAAATGCGCGGCATC 345
101 LeuGluGlyLysIleValSerValGlnTrpProAlaSerValAr 117
|||||
346 TTGCTCGAGACACCGCTGCTGCTCCACTGCGCAGCGTGGTGGC 395
117 gThrIleLeuGlySerAspTrpLeuSerAsnValHisGlyValGlnHisL 134
|||||
396 CACCATCTGGATCTGCTGCTCTTCACTCACTGACGACTCTCGCACA 445
134 ysaAsnLysLysAlaIleMetArgAlaPheSerArgAspAlaLeuGlu 150
|||||
446 AGCAGCGCAAGAGTGTATGTGGGGCTTCACCGCGAGGACCTCGAA 495
151 HisTrpIleProValIleGlnGlnValLysSerAlaIleGlnGluTr 167
|||||
496 TGCCTACGTGCGGTATCACCAGGAAGTGGGCGAGCCGTGGAGCAGTG 545
167 pLeuGlnLysAspSerCys.....ValLeuValTrpProGluM 180
|||||
546 GCTG.....AGCTGCGCGAGCGCGGCTCTGCTACCCGAGG 586
180 eLysLysLeuMetPheArgIleAlaMetArgIleLeuLeuGlyPheGlu 196
|||||
587 TGAAGCGCTCATGTCCGAATCCGATCGCATCTACCTGCGGTGGAA 636
197 ProGlnIleLysThrAsp.....GlnGlnIleLeuValGlnAla 210
|||||
637 CCC...CAACTGGCGGGGAGCGGGGACTCCGAGCAGCAGCTGTGGAGGC 683
210 aPheGlnGlnMetIleLysAsnLeuPheSerLeuProIleAspValProp 227
|||||
684 CTTCGAGAGAATGACCCGCAATCTCTCTGCTCCATCGACGCTGCCCT 732
227 heSerGlyLeuTrpArgGlyLeuArgAlaArgAsnPheIleHisSerLys 243
|||||
733 TCACGCGGCTGTACCGGGGATGAAGGCGCGGACATCATCGACGCGCG 782
244 IleGlnGlnAsnIleArgLysLysIleGlnAsp...AspAspAsnGlnAs 259
|||||
783 ATCGAGCGCAACATTCGCGCGCAAGATCGCGGCTGCGGCGCGCGAGCC 832
259 nGlnGlnLysTrpLysAspAlaLeuGlnIleLeuIleGlnAsnSerArg 276
|||||
833 GGGCGAGGGCTGCAAGACGCGCTGCAAGCTGTGATCGAGACACGAGG 882
276 rGSerAspGlnPropheSerLeuGlnAlaMetLysGlnAlaIleThrGlu 292
|||||
883 AGAGGGGAGACGGCTGACATGACAGCGCACTAAAGCAATCTCAACCGAA 932
293 LeuLeuPheGlyGlyHisGlnValThrThrAlaSerThrAlaThrSerLeuVal 309
|||||
933 CTCTCTTTGGAGACAGCAAGACACGCGCGCTGCGACCGACATCTCGAT 982
309 lMetPheLeuGlyLeuAsnTrpGlnValValGlnLysValArgGlnGluVal 326
|||||
983 CACTTACCTGGGCTTACCCACCATGTTCTCCAGAAAGTGGAGAAAGAGC 1032
326 aLingGlnLysValGlnMetGlyMetTrpTrpProGlyLysGlyLeuSer 342
|||||
1033 TGAAGAGTAAGGTTTACTTTCGAAGAGCATCAAGCAACAAATTGGAC 1082
343 MetGlnLeuLeuAspGlnLeuLysTrpThrGlyCysValIleLysGlnTrp 359
|||||
1083 ATGGAATTTTGGAAACAATTAAATACATCGCGTGTCTTTAAGAGAC 1132
359 rLeuArgIleAsnProproValProGlyGlyPheArgValAlaLeuLysTr 376
|||||
1133 CCTTGCAGCTGAATCCCGCAGTTCACAGAGGGTTTCGGGTTGCTGTGAAGA 1182

```

```

376 hrPheGlnLeuLeuAsnGlyTrpGlnIleProLysGlyTyrAsnValIleTyr 392
|||||
1183 CTTTGAATTTAAATGATACCAAGATTCGCCAAGCGCTGGAATTTATCTAC 1232
393 SerIleCysAspThrHisAspValAlaAspValPheProAsnLysGlnL 409
|||||
1233 AGTATCTGTGATACATCATGATGTGGCAGAGATCTTCACCAACAGGAAGA 1282
409 upheGlnProGlnIleArgPheMetSerLysGlyLeuGlnAspGlySerArgP 426
|||||
1283 ATTTAATCTTGACCAATTCATGCTGCTCACCACAGAGATGATCCAGCT 1332
426 heAsnTrpIleProPheGlyGlyGlySerArgMetCysValGlyLysGlu 442
|||||
1333 TCAGCTTCATTCATTTGAGAGAGCGCTTACGAGCTGTGTAGCAAAAGAA 1382
443 PheAlaIleValLeuLeuLysIlePheLeuValGlnLeuThrGlnHisCys 459
|||||
1383 TTTCGAAAATTTCTTCAAAATATTACAGTGAAGTGGAGCGCAGCATTTG 1432
459 saSnTrpIleLeuSerAsnGlyProProThrMetLysThrGlyProThrI 476
|||||
1433 TGACTGCGAGCTTCTTAATGAGACTCTCTCAATGAAGAACAGTCCACCG 1482
476 leryProValAspAsnLeuProThrLysPheThrSerTyr 489
|||||
1483 TGTATCTGTGGACAATCTCCGCAAGATTCACCATTTTC 1523
seq_name: /SIDSL/gcdata/geneseq/geneseqn/NA1998.DAT:AAV12216
seq_documentation_block:
ID AAV12216 standard; DNA; 4164 BP.
XX
AC AAV12216:
XX
DT 22-JUN-1998 (first entry)
XX
DE Mouse retinoid metabolizing protein mp450RAI genomic DNA.
XX
KW Retinoid metabolizing protein; P450RAI; retinoid oxidase;
KW retinoic acid; mouse; inhibitor; antisense; cancer;
KW actinic keratosis; oral leukoplakia; head tumour; neck tumour;
KW non-small cell lung carcinoma; basal cell carcinoma;
KW acute promyelocytic leukaemia; skin cancer; acne; psoriasis;
KW ichthyosis; therapy; diagnosis; screening; cytochrome P450; ds.
XX
OS Mus musculus.
XX
PN M09749815-A1.
XX
PD 31-DEC-1997.
XX
PF 23-JUN-1997; 97MO-CA00440.
XX
PR 01-OCT-1996; 96US-0724466.
XX
PR 21-JUN-1996; 96US-0667546.
XX
PA (TOOH ) UNIT QUEENS KINGSTON.
XX
PI Beckett BR, Jones G, Petkovich PM, White JA;
XX
DR WPI; 1998-077178/07.
XX
PT Retinoid metabolizing protein - useful to develop products to treat,
PT e.g. cancer, actinic keratosis, oral leukoplakia, acne, psoriasis or
PT Ichthyosis
XX
PS Disclosure; Page 69-70; 110pp; English.
XX
CC This sequence comprises a genomic clone of a novel mouse retinoid
CC metabolizing protein designated mp450RAI. A cDNA clone (see
CC AAV12205) that includes a coding sequence for the full-length
CC mp450RAI protein (see AAV44161) is also provided. The encoded

```

CC protein is a retinoid oxidase that has the ability to hydroxylate
 CC retinoic acid at the 4 position of the beta-ione ring and is
 CC inducible in epithelial cells exposed to retinoic acid. Zebrafish,
 CC human and mouse p450RAI cDNA clones (see AAV1203-05) are claimed.
 CC They can be used in the recombinant production of p450RAI.
 CC Antisense sequences can be used in a claimed method for inhibiting
 CC retinoic acid hydroxylation for the treatment of cancer, actinic
 CC keratosis, oral leukoplakia, secondary tumours of the head and/or
 CC neck, non-small cell lung carcinomas, basal cell carcinomas, acute
 CC promyelocytic leukemia, skin cancer, and premalignancy associated
 CC with actinic keratosis, acne, psoriasis and/or ichthyosis. Promoter
 CC sequences (see AAV12206-08) for the p450RAI genes are also claimed.
 XX
 SQ Sequence 4164 BP; 997 A; 975 C; 1071 G; 1121 T; 0 other;

alignment_scores:

Quality: 1169.00 Length: 1055
 Ratio: 3.005 Gaps: 16
 Percent Similarity: 36.872 Percent Identity: 29.858

alignment_block:

US-09-668-482-2 x AAV12216

Align seg 1/1 to: AAV12216 from: 1 to: 4164

34 ArgArgValAspProAsnCysArgSerProLeuProPro..... 46
 ||| ||||| ::| ||||| |||||
 25 AGAGGGGTAGACGAGACGATGCTCCCGGGCCCTCATGCGGGTTG 74
 47 .GlyThrMetGlyLeuProPheIleGlyGluThrLeuGlnLeuIleLeu 63
 ||| |||||
 75 CGGTCTCTCTCTCACT.....CCCTCTC 100
 63 LArgArgLysPheLeuArgMetLysArgGlnLysTyrGlyCysIleTyr 79
 ||| ||||| ::| ||||| |||||
 101 ACGCGAGGAAGTTTTCGACATGAAGCCGAGAAATACGGCTTCATCTAC 150
 80 LysThrHisLeuPheGlyAsnProThrValAlaValMetGlyAlaAspAs 96
 ||| ||||| ::| ||||| |||||
 151 AAGACCACTCTTTGGGGGCGCCACGGTGGGTGATGGCGCGGATTA 200
 96 nValArgGlnIleLeuGlyGluHisLysLeuValSerValGlnTrpP 113
 ||| ||||| ::| ||||| |||||
 201 TGTGGCGGCATCTTCTGCGAGACACCGGTTGCTGCTGCACTGGC 250
 113 roLAserValArgThrIleLeuGlySerAspThrLeuSerAsnValHis 129
 ||| ||||| ::| ||||| |||||
 251 CCGCTCGGGGCGCACCATCTGGGCGCTGCTCTCAACCTGCAC 300
 130 GlyValGlnHisLysAsnLysLys..... 138
 ::| ||||| ::| |||||
 301 GATTCCTCGCAAGCAGCAAGAGGTGAGGGTGAGCTGGCACTCCT 350
 138 138
 351 TGGCTGGCAGGAGACCTCATCTATGGCTTGGTTCAGGAAATAGAAAT 400
 138 138
 401 GCGGGGAGGAGGCTAGTCTATGTGGGGAGCAGACCTCTCTATCT 450
 138 138
 451 GAGATCCACTTTAGCTTTCTGCTAGACAGTGGTTAGTCTGGGGGA 500
 138 138
 501 CTGAATTTCTGAAGGGTACTCGAAAGCGAAGGGGGGGGCTGAGG 550
 139Ala,Il 140
 551 GAAAGTAGAGGATTTGTAACACTCTGCTCTGGGGGGTCTCAGGTGAT 600

140 eMetArgAlaPheSerArgAspAlaLeuGlnHisTyrIleProValIleG 157
 ||| ||||| ::| ||||| |||||
 601 TATGCAAGCCCTCAGCCGAGGACGAGTCTAGTCTAGTCCCGGTATCG 650
 157 lngIngluValLysSerAlaIleGlnIleuTrpLeuGlnLysAspSerCys 173
 ::| ||||| ::| |||||
 651 CTAGAGAGTCACAGAGTTGCTGAGACAGTGCTA.....ACGTGC 691
 174ValLeuValTyrProGluMetLysLysLeuMetPheAr 186
 ::| ||||| ::| |||||
 692 GCGAGCGCGGCTCCTCGTCTACCCCGAGGTGAAGCGCTCATGTTCG 741
 186 gIleAlaMetArgIleLeuLeuGlyPheGluProGluGlnIle.....L 201
 ||| ||||| ::| |||||
 742 CATGCGCATGGCATCTCTCGGTGCGAGCCGGTCCAGCGGGGGCG 791
 201 ySThrAspGluGlnGluLeuValGluAlaPheGluGluMetIleLysAsn 217
 ||| ||||| ::| |||||
 792 GCGAGGACGAGCAGCAGCTCGTGGAGGCTTTCGAGGAGATGACCCGCAAT 841
 218 LeuPheSerLeuProIleAspValProPheSerGlyLeuTyrArg.... 232
 ||| ||||| ::| |||||
 842 CTCTTCTCTTCCCATTTGAGTGCCCTTTAGCGGCTGTACCGGGTAA 891
 232 232
 892 GCGCGTTTGGCGAGTCGGAGTAGGGGAACGACGTGGGCATCCGCTCA 941
 233Gly,LeuArgAlaArgAsnPhelIleH 241
 ||| ||||| ::| |||||
 942 CCGCAGCCTCTCTCCCGCTCAGGCGGTGAAGCGCGGAACCTTATACA 991
 241 sSerLysIleGluGlnLysAsnIleArgLysIleGlnAsp...AspAspA 257
 ||| ||||| ::| |||||
 992 CGCGGCACTCGAGGAAACATTCGCGCAAGATCCGGGCTTCAGGCTA 1041
 257 snGluAsnGluGlnLysTyrLysAspAlaLeuGlnLeuIleGluAsn 273
 ::| ||||| ::| |||||
 1042 CAGAGCCGAGTGGGGGTTCGAAGAGCGGCTGACGCTCTATTGACAC 1091
 274 SerArgArgSerAspGluProPheSerLeuGln..... 284
 ||| ||||| ::| |||||
 1092 TCGTGGGAGAGGGAGAGAGGCTGATATGACGATGAGAACAAATTCAA 1141
 284 284
 1142 AAGGTCCCAAGGGCGGGGAGTGCTGTGACTTTCAGACACACTTTCTG 1191
 284 284
 1192 GGGTCCCAAGCCCTGTCAAGGCCAGTACTTCCAAGTGGGGCGGA 1241
 284 284
 1242 TGCTAGGTCTAGAGCTTTCAACCTGTGGGTGTCGACCCCTTCAGCAGC 1291
 284 284
 1292 CAACAACCCCTTTCAGAAAGGTGCGCTAAGAGCATCGATATCCGATAT 1341
 284 284
 1342 TTACATCAGAAACATACAGTAGCAAAATTACCTTATGAGTAGAACAC 1391
 284 284
 1392 AAAGATATTTTATCGTTGGGGGTCAACCAACACAGAGAACCGTATTA 1441
 284 284
 1442 AGGGTGCAATTGGTCTAGAGAGCTGTGAAGGGGTGGCTGAGCAATGG 1491

```

284 ..... 284
1492 GAAGATCCCAAGTCAAGGCGACGCTCATCTCAAAAGTTAAAGCG 1541
284 ..... 284
1512 AAGACGAGATTAAAGGAGTTTGGCTTTTGTGTGCTTTGACTT 1591
284 ..... 284
1592 CTATGACAAAAAGGATTTACCTTGAAGTCTCCGTGCAATATTC 1641
284 ..... 284
1642 GGTGAGTCTTTGTACAGTGTCTATAATGCACTCAGATCTGTATA 1691
285 ..... 285
1692 TTCCGTTTTTATCCCTTAGCCTAAACATCGTCACAGAGCTCTCT 1741
295 eglYglYHisgluThrAlaSerThrAlaThrSerLeuValMetPheL 312
1742 TGGTGTATCAAACTACAGCAGCTGCTGCGACGCTCACTGATCACT 1791
312 eugLYleuAsnThrGluValGlnLysValArgGluGluValGlnL 328
1792 TAGGACTCTACCCACATGTCTCTCCAGAAAGTTGAGAGAGATAA 1841
329 LysVal ..... 330
1842 AAGGTAGAGATTTCTAGAGGTCCCATTTGCCTAGACATCTCTAT 1891
330 ..... 330
1892 TAACCAACACACACACCCACCTGATATAGTTTGCCTGATACACC 1941
330 ..... 330
1942 TACTATGACAGTGAAGATCTGAGAGCTAGTGGAGCTGGGGAGAG 1991
330 ..... 330
1992 TCCACCTCGTAATTTAAAAAGCAGTTGTTGACTGGGCTCTCTTG 2041
330 ..... 330
2042 GGCAGAAATTGACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2091
330 ..... 330
2092 TCCACCAACACACACACCTTTATAGAGCAAGTTCTCTCTTCC 2141
330 ..... 330
2142 CTGACCAAGACATGAATATGTAGATTAGACCAATAGCTGATCAGG 2191
331 ..... 339
2192 GCAGTGTGGTAGGGCTCAGGGTATGACCTTTATATACCTGTATA 2241
339 ..... 339
2242 ACATGTCTGATATATGGTTTAGGCTGAGGAAGTGTGAAAGAGGCC 2291
339 ..... 339
2292 ATCAGGCCATCAGCTCTTCCCTTTATCTCTCCATCCAGAGCCTTC 2341
339 ..... 339
2342 AGGTTTAGTTAACAGGTGAGTCTCTGGGCTGACTTTTGTGAGTG 2391
339 ..... 339

```

```

2392 CCCAGGATCCATCACTCACTTTTATCTGTTTCCATAGGGCTTACT 2441
340 ..... 349
2442 GCAAGAGCAATCAAGACAAACAGTTAGACATGGAACCTTGGCACAG 2491
350 LysThrThrGlyCysValIleLysgluThrIleuArgIleAsnPro 366
2492 AATACACTGGGTGTGTCAATTAAAGAGACCCCTGCATGGAATCTC 2541
366 lProGlyGlyPheArgValAlaLeuLysThrPheGlu ..... 378
2542 TCCAGAGGGTTTCGGGTTCCTCTGAGAGCTTTGAGCTGAATGTG 2591
378 ..... 378
2592 CACCTCTGTCCCCACCCACCCCTCGTCCAGCTCCACTGTCTATGC 2641
378 ..... 378
2642 TGTGAGCATCAGCTGCCAGACAGAGTGGCTCAGTCCCTTGACAG 2691
378 ..... 378
2692 CTGCCTCTATGCTAGTGGAAACCAATTGCTCTCTCTTAATGCCAT 2741
378 ..... 378
2742 CCAGCTAGTATGACTTTTGTGTTGTCAAAGCTCAGGCGGAGTTGTC 2791
379 ..... 386
2792 AATCTTAGATTTTGTGTTTATTAACAGGATACCAATCCCAAG 2841
387 GlyTrpAsnValIleTyrSerIleCysAspThrHisAspValAla 403
2842 GGTGAGATGTTATTTACATATCTGTGACACCAACCATGTGGCAGAT 2891
403 lPheProAsnLysGluGluPheGlnProGluArgPheMetSerLys 420
2892 CTTCACTAACAGAGGAATTTAATCCGACCGCTTCTTAAGATTTAC 2941
420 eugLysAspGlySerArgPheAsnTyrIleProPheGlyGlySer 436
2942 CAGAGATGCTTCCGGTTCAGCTTCAATTCATTTGGAGAGGCTTCG 2991
437 MetCysValGlyLysGluPheAlaLysValIleuLysIlePheLe 453
2992 AGCTGTAGCCAAAGATTGCAAAAATTTCTTAAAGATATTACAGT 3041
453 lGluLeuThrGlnHisCysAsnTrpIleLeuSerAsnGlyProPro 470
3042 GAGCTGCTAGGACGCTGTGATTTGCAAGCTTTCTAAATGACCTCT 3091
470 eLysThrGlyProThrIleTyrProValAspAsnLeuProThrLys 486
3092 TGAAGCAACACCCACGCTGTACCTGTGACAAATCTCCCTGCAAG 3141
487 ThrSerTyr 489
3142 ACCCACTTC 3150

```

seq_name: /SIDS1/gcdata/geneseq/geneseqn/NA198.DAT:AAV12214

seq_documentation_block:

ID AAV12214 standard: DNA; 2677 BP.

XX AAV12214;

AC

XX

XX 22-JUN-1998 (first entry)

DT

XX

DE Human retinoid metabolising protein hp450RAI genomic DNA.

XX Retinoid metabolising protein; P450RAI; retinoid oxidase;
 KM retinoic acid; human; inhibitor; antisense; cancer;
 KM actinic keratosis; oral leukoplakia; head tumour; neck tumour;
 KM non-small cell lung carcinoma; basal cell carcinoma;
 KM acute promyelocytic leukaemia; skin cancer; acne; psoriasis;
 KM ichthyosis; therapy; diagnosis; screening; cytochrome P450; ds.
 OS Homo sapiens.
 XX
 XX W09749815-A1.
 PN
 XX
 PD 31-DEC-1997.
 XX
 XX 23-JUN-1997; 97WO-CA00440.
 PF
 XX 01-OCT-1996; 96US-0724466.
 PR 21-JUN-1996; 96US-0667546.
 XX
 XX (TOOH) UNIT QUEENS KINGSTON.
 PA
 PI Beckett BR, Jones G, Petkovich PM, White JA;
 DR WPI; 1998-077178/07.
 XX
 PT Retinoid metabolising protein - useful to develop products to treat,
 PT e.g. cancer, actinic keratosis, oral leukoplakia, acne, psoriasis or
 PT ichthyosis
 PS Disclosure; Page 68; 110pp; English.
 XX
 XX This sequence comprises a genomic clone of a novel human retinoid
 CC metabolising protein designated hp450RAI. Another genomic clone
 CC (see AAV12215) is also provided, as well as a cDNA clone (see AAV12204)
 CC that includes a coding sequence for the full-length hp450RAI
 CC protein (see AAW4160). The encoded protein is a retinoid oxidase
 CC that has the ability to hydroxylate retinoic acid at the 4 position
 CC of the beta-ionone ring and is inducible in epithelial cells
 CC exposed to retinoic acid. The hp450RAI gene has been localised to
 CC 10q23-24. Zebrafish, human and mouse P450RAI cDNA clones (see
 CC AAV12203-05) are claimed. They can be used in the recombinant
 CC production of P450RAI. Antisense nucleic acids can be used in a
 CC claimed method for inhibiting retinoic acid hydroxylation for the
 CC treatment of cancer, actinic keratosis, oral leukoplakia, secondary
 CC tumours of the head and/or neck, non-small cell lung carcinomas,
 CC basal cell carcinomas, acute promyelocytic leukaemia, skin cancer,
 CC and premalignancy associated with actinic keratosis, acne,
 CC psoriasis and/or ichthyosis. Promoter sequences (see AAV12206-08)
 CC for the P450RAI genes are also claimed.
 XX
 XX Sequence 2677 BP; 489 A; 775 C; 868 G; 511 T; 34 other;
 SQ
 alignment_scores:
 Quality: 792.00 Length: 673
 Ratio: 2.891 Gaps: 13
 Percent Similarity: 40.713 Percent Identity: 31.947
 alignment_block:
 US-09-668-482-2 x AAV12214 ..
 Align seg 1/1 to: AAV12214 from: 1 to: 2677
 1 MetGlyLeuTyrThrLeuMetValThrPheLeuCysThrIleValLeuPr 17
 ||||||| :|||:|||||: |||||||:|||||:|||||
 676 ATGGGGCTCCCGCGCTGCTGCGCAGTGGCGCTTCACCTTCGTCGTCGCC 725
 17 oValleuLeuPheLeuAlaIleValLysLeuTrpGluMetLeuMetIleA 34
 ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 726 GCGGCGCTTCCTGCGCTGCGATCAAGCTTCGGACCTGACTCGCGTGA 775
 34 rGArGValaSProAnScYArGSePrOleuProGluLyrMetGly 50
 ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||

776 GCGGCGGAGCCGACGTTGTGCGCTCCCATTTGCCCGCGGACATATSGGS 825
 51 LeuPProPheIleGlyGluThrLeuGlnIleLeuGlnAlaArgLysP 67
 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 826 TTCCCTCTCTTGGGGAAACCTTGACATGATCTACAGTAAGGAGAGG 875
 67 heLeuArg..... 69
 |||
 876 TGGGGCGGACAGCGCTGCTCCCGGAGCCCGCGGCTCTGGGCTTCT 925
 69 69
 926 GGTGAAGTCGGGTAGCGCCCCCGGAGCATGCTATTGCGGTAGGAG 975
 69 69
 976 CAGGCTGCGGGAGCGCGCGCTCCCGGMYKMSCTCAGSCRCMKK 1025
 70 MetLysArg 72
 |||||||
 1026 TMMCTCCGCTTCTCCCGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1075
 73 GlnLysTyrGlyCysIleTyrLysThrHisLeuPheGlyAsnProThyA 89
 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 1076 AGGAATACGCGCTTCATCTAACAAGCGCATCTGTCGGCGGCCACCGT 1125
 89 larGValMetGlyAlaAspAsnValArgGlnIleLeuGlnGlyLysHst 106
 |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 1126 ACGGCTGATGGCGCGGACAAATGCGCGCATCTTGTGCGAGAGCACC 1175
 106 yLseuValSerValGlnTrpProAlaSerValArgThrIleLeuGlySer 122
 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 1176 GCGTGTGTGCTGCTCACTGGCGCGCTCGTGCGACCAATCTGGGATCT 1225
 123 AspThrLeuSerAsnValHisGlyValGlnHisLysAsnLysLys.. 138
 |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 1226 GCGTGCCTCTTAACCTGCGACGACTCTCGCACAGCAGCAAGAGT 1275
 138 138
 1276 GGGGCGAGGAGCGGCGCTGCGACAGGAGGAGGAGGAGGAGGAGG 1325
 138 138
 1326 GAATTCGGGTATGATGCTAGCGCGGCTAGCAGCTTAGGTGGGCT 1375
 138 138
 1376 AGGACCTCTGCCAGCTCAGGTTAGCTTCCAGCTCGGAGAGTGCAT 1425
 138 138
 1426 GTGCTGCGACGACCTGGGGGTGTCTGAAGGAGCGGCGGTAGACGAG 1475
 139 Alalle 140
 1476 GGGCGATGAGGCTTTAAAGCTGTCCCTCTCGGAGCTCAGGTGAT 1525
 141 MetArgAlaPheSerArgAspAlaLeuGlnHisTyrIleProValIleG 157
 |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 1526 ATGGGCGCTTCAGCGCGAGGCACTCGAATGCTACCTGCGCGGTGATAC 1575
 157 nGlnGluValLysSerAlaIleGlnIleLysLeuGlnLysAspSerCys. 173
 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 1576 CGAGAGAGTGGCGAGCGCTCGAGAGAGTGGCTG..... AGCTGCG 1616
 174 ValLeuValTyrProGluMetLysLysLeuMetPheArg 166
 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 1617 GCGAGCGGCGCTCTGCTGCTACCCCGAGGAGGAGCGCCATGTTCCGA 1666
 187 IleLeuMetArgIleLeuGlnGlyPheGluProGluLysIleLysTrpAs 203
 |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 1667 ATCGCATGCGCATCTTACTGCGCTGCGAACC... CAACCTGGCGGCGGA 1713

```

203 P.....GluGlnGluLeuValGluAlaPheGluGluLeuTyrAsp 217
      |||||.....:|||||
1714 CGGGAGACTCGAGCAGCAGCAGCTGTGGAGGCGCTTGAGAAATGACCCGCA 1763
      |||||.....:|||||
217 sntLeuPheSerLeuProIleAspValProPheSerGlyLeuTyrArg... 232
      |||||.....:|||||
1764 ATCTCTCTGCTGCTCCATCGACCTGCTCCCTTACGCGGGCTGTACCGGTA 1813
      |||||.....:|||||
232 ..... 232
1814 AGGGCGGCAACAGGGCTGCGGACTAGGGCGCGGACCTGGGCGCTGCT 1863
233 .....Gly..LeuArgAlaArgAsnPheIle 240
      |||.....:|||||
1864 CACCGCGCGCGCTCTCTGCGCTCAGGCGCATGAAAGCGCGGAACCTCAT 1913
241 HisSerLysIleGluGluAsnIleArgLysLysIleGlnAsp...AspAs 256
      |||||.....:|||||
1914 CACGCGGCGCATCGAGCAGAACATTCGCGCAAGATCTGCGGGCTGCGGCG 1963
256 pAsnGluAsnGluGlnLysTyrLysAspAlaLeuGlnLeuLeuIleGlu 273
      |||||.....:|||||
1964 ATCCGAGGGCGGCGCAGGGCTGCAGAAAGCGCGCTGACGCTTGTATCGAGC 2013
273 snSer..... 274
      |||||
2014 ACCTCTGGGAGAGAGGAGAGCGGCTGGACATGCAGGTAGTAGCAGCTTC 2063
274 ..... 274
2064 AGACCAAGCACTGCGAGTTTGCTCCCTGCTTCCAAAGCGCTGTCC 2113
274 ..... 274
2114 TGGGGCCCCCAAGAGCGGCGCTGGGGCCACGCTTCTGGAGTGGGCGGC 2163
275 .....ArgArgSerAspGluPro... 280
      |||||.....:|||||
2164 CGGCTCAGACTACAGCTATGATCCGGAAGAGCTGAGACACCCGGT 2213
280 ..... 280
2214 CAGGAGAGCTGGGGAAGGGGCTGCGGAGCAACTGGAGCACTCCCTAGC 2263
280 ..... 280
2264 CTTCACAGCTTCAAGGGAAGTTGGAATTTGCAAAAATGTTAATAA 2313
280 ..... 280
2314 GAACCTTGGATTTAATAAATAAGACTTAACTCAGAGAGTTCCGGT 2363
280 ..... 280
2364 AGRGCGGGTCTGCTACTCGGCTTACTGCTCCAGCTGAACATAAGGACGTT 2413
280 ..... 280
2414 GCATTTGTTAAAGATATGCTTCTTCTGACTTCTGTCAGCAAAACAT 2463
280 ..... 280
2464 TTAGCCCTTAGTCTTCCCTTCAGAACTCTCAGTTCGATTCGTAGTAAT 2513
281 .....PheSerIle 283
      |||||
2514 CTTCTGTCAAAACGAGCAGACTTGTGAGAAATGGGTCTCACTCTAT 2563
283 uGln...AlaMetLysGluAlaAlaThrGluLeuLeuPheGlyGlyHisG 299
      |||||.....:|||||
2564 TCTTAGGCACTAAAGCAATCTTCAACGAACTCTCTTTGGAGGACAGCA 2613

```

```

299 uThrThrAlaSerThrAlaThrSerLeuValMetPheLeuGlyLeuAsnT 316
      |||||.....:|||||
2614 AACCAAGGCGCAGTCAGCAGCATCTGTATCATCTTACTGGGCTCTAAC 2663
      |||||.....:|||||
316 hrGluValValGln 320
      |||||.....:|||||
2664 CACATGTTCTCCAG 2677
seq_name: /STD1/gcgdata/geneseq/geneseqn/NA1997.DAT.AAT85306
seq_documentation_block:
ID AAT85306 standard; cDNA to mRNA; 1608 BP.
XX
AC AAT85306;
XX
DT 14-APR-1998 (first entry)
XX
DE Arabidopsis thaliana cytochrome P450-type hydroxylase cDNA.
XX
KW Cytochrome P450-type hydroxylase; identification; brassinosteroid;
KW brassinosteroid inhibitor; modified plant; recombinant production;
KW testosterone; ds.
XX
OS Arabidopsis thaliana.
XX
FH Key Location/Qualifiers
FT CDS 48..1467
FT /tag=a
FT /product= cytochrome_P450-type_hydroxylase
XX
XX WO9735986-A1.
XX
XX 02-OCT-1997.
XX
XX 27-MAR-1997; 97WO-EP01586.
XX
XX 27-MAR-1996; 96US-0622166.
XX
XX (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX
XX Altmann T, Koncz C, Mathur J, Szekeres MA;
XX
XX WPI; 1997-489649/45.
XX
XX P-PSDB; AAM27153.
XX
XX New isolated plant cytochrome P450-type hydroxylase gene - used to
XX identify substances acting as brassino-steroid(s) or brassinosteroid
XX inhibitors for the production of modified plants
XX
XX PT
XX PS Claim 1; Pages 44-46; 77pp; English.
XX
XX The present sequence encodes Arabidopsis thaliana cytochrome
XX P450-type hydroxylase. The hydroxylase can be used to identify
XX brassinosteroids or brassinosteroid inhibitors, useful to produce
XX plants with modified physiological and/or phenotypic
XX characteristics. The modified plants may show, e.g. stimulation of
XX growth, increased cell elongation, increased wood production,
XX accelerated seed germination at low temperatures, an increase in
XX dry weight, repressed anthocyanin production during growth in light
XX and/or inhibited de-etiolation which is induced, e.g. by cytokinin,
XX in the dark or an increase in stress tolerance. The hydroxylase or
XX its coding sequence can also be used for the recombinant production
XX of compounds, e.g. testosterone.
XX
XX Sequence 1608 BP; 382 A; 374 C; 374 G; 478 T; 0 other;
XX
alignment_scores:
Quality: 552.50 Length: 485
Ratio: 1.811 Gaps: 14
Percent Similarity: 62.887 Percent Identity: 29.897
alignment_block:

```


XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 22-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134370.
PR 19-MAY-1999; 99US-0134768.
PR 20-MAY-1999; 99US-0134941.
PR 21-MAY-1999; 99US-0135124.
PR 24-MAY-1999; 99US-0135353.
PR 25-MAY-1999; 99US-0135629.
PR 27-MAY-1999; 99US-0136021.
PR 28-MAY-1999; 99US-0136392.
PR 01-JUN-1999; 99US-0136782.
PR 03-JUN-1999; 99US-0137222.
PR 04-JUN-1999; 99US-0137528.
PR 07-JUN-1999; 99US-0137502.
PR 08-JUN-1999; 99US-0137724.
PR 10-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0138847.
PR 16-JUN-1999; 99US-0139119.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 24-JUN-1999; 99US-0140354.
PR 28-JUN-1999; 99US-0140699.
PR 29-JUN-1999; 99US-0140823.
PR 30-JUN-1999; 99US-0140991.
PR 01-JUL-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 08-JUL-1999; 99US-0142390.
PR 09-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.

PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145224.
PR 27-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145919.
PR 02-AUG-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 09-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 10-AUG-1999; 99US-0147935.
PR 11-AUG-1999; 99US-0148171.
PR 12-AUG-1999; 99US-0148319.
PR 13-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148665.
PR 16-AUG-1999; 99US-0148684.
PR 17-AUG-1999; 99US-0149368.
PR 18-AUG-1999; 99US-0149175.
PR 20-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 23-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.

```
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161933.
PR 29-OCT-1999; 99US-0162142.

alignment_scores:
Quality: 552.50 Length: 485
Ratio: 1.817 Gaps: 14
Percent Similarity: 62.887 Percent Identity: 29.897
```

alignment_block:
US-09-668-482-2 x AAC48157 ..

Align seg 1/1 to: AAC48157 from: 1 to: 1646

```
19 LeuLeuPheLeuAlaValLysLeuTrpGluMetLeuMetLeuArg 35
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
77 CTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCGCG 126
35 GValAspProAsnCysArgSerProLeuProGlyThrMetGlyLeuP 52
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
127 TACACGTTACCGT...CGATGGGTCCTCCCGGGAAGCCTTGCTCC 173
52 roPheIleGlyIuThrLeuGluLeuLeu.....LeuGln 63
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
174 CTCTGATTAGAGAGACTTTTCAGCTGATCGAGCTTACAAAACAGAGAAC 223
64 ArgArgLysPheLeuArgMetLysArgGlnLysTyrgLysIleTyrl 80
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
224 CGTAGAGCCTTCATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 273
80 sThrHisLeuPheGlyAsnProThrValArgValMetGlyAlaAspAsn 97
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
274 GACGCATCTTTTGGTGAACGAGCATTTTCTCAGCTGACCGGAAACGA 323
97 aArgGlnIleLeuGluGlyIuHisLysLeuValSerValGlnTrpPro 113
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
324 ACCGGTTTGTCTTCAGAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 373
114 AlaSerValArgThrIleLeuGlySerAspThrLeuSerAsnValHisG 130
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
374 GCTTCATTTGTAACCTTTGGGGAACACTCTCTGCTTCTATGAAAG 423
```

```
130 yValGlnHisLysAsnLysLysValAlaIleMetArgLaphSerArg 147
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
424 TTCCTTGGCATAAA...CGTATGCACTCTCTCACCATGACCTTGTATAT 470
147 sPalLeu.....GluHisTyrlleProValIleGlnGlnGluValLys 161
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
471 CTTCATCATTTAAGACCATCTCATGATATGACCGGTAGTCCGG 520
162 SerAlaIleGlnGluThrLeuGlnLysAspSerCysValLeuValTyPr 178
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
521 TTATATCTGATTCCTGG.....TCTCTGCTGCTCTCTCATGGA 561
178 oGluMetLysLysLeuMetPheArgIleAlaMetArgIleLeuGlyP 195
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
562 AGAAGCCAAAAGATATACGTTGACGTACGTAAGCAAGCTGATGAGAGCT 611
195 heGluPro.....GluGlnIleLysThrAspGluGlnGluLeu 207
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
612 TTGATCCAGGGAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 661
208 ValGluAlaPheGluGluMetIleLysAsnLeuPheSerLeuProIleAs 224
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
662 ATCGAAGGCTC.....TCTCTCTCTCTCTCC 690
224 pValProPheSerGlyLeuTyArg...GlyLeuArgAlaArgAspPhe 240
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
691 TCTC...TTCCTCACCACTTACCGCAAGGCATCCAAAGCGGAGAGAG 737
240 leHisSerLysIleGluGluAsnIleArgLysLysIleGlnAspAsp 256
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
738 TGGCGAGCGCTTGACGGTGCTGTATGATAAAG...AGGGAGGAGGAG 784
257 AsnGluAsnGluGlnLysTyrlLysAspAlaLeuGlnLeuLeuIleGln 273
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
785 GAAGAAAGAGCGAGAGAGAAAGAAAGATATGCTTCGCGCTTGGCTT... 829
273 nSerArgArgSerAspGluProPheSerLeuGlnAlaMetLysGluAla 290
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
830 .....GCGCGGATGATGATGATTTTCCGATGAAGATTTGATGACTCT 872
290 IatHrGluLeuLeuPheGlyGlyHisGluThrAlaSerThrAlaThr 306
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
873 TGGTGTCTTACTTGTCCGCGGTATATGAACAACCTCCACGATCATGACT 922
307 SerLeuValMetPheLeuGlyLeuAsnThrGluValGlnLysValAr 323
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
923 CTGCGCCGTAATTTCTCACCAGACCTCTTAGCTCTCTCAACTCA 972
323 gGluGluValGlnGluLysValGluMetGlyMetTyrlThrProGlyLys 340
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
973 GGAAGAG...CATGAAAAGATTAG...GCAATGAAGAGATTGCGTAT 1016
340 LysLeuSerMetGluLeuLeuAspGlnLeuLysTyrlThrGlyCysValIle 356
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1017 GCTTGAATGAGATTTACAAAGTCAATGCCATTCACACATATGTTGCTT 1066
357 LysGluThrLeuArgIleAsnProProValProGlyLysPheArgValAl 373
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1067 AATGAGAGCTGACGAGTGCGCTACATCATCGCGGCTTTTCAGACCTGC 1116
1117 AATGAGGATTTGAGATCAAAAGTTTAAATTCCAAAGAGGAGGAAG 1166
390 aIleTyrlSerIleCysAspThrHisAspValAlaAspValPheProAsn 406
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1167 TATCTCATCTGTTAGACCGGTTTATAGACCAAAACCACTTAAAGAT 1216
407 LysGluGluPheGlnProGluArgPheMetSerLysGlyLeuGlnIleAsp 423
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1217 GCTGCACCTTTCACCCCTTGAGATGCGAGAGCAATCGTAAACGAGAG 1266
```

```
423 ySerArgPheasnTyrlleProphleGlyGlySerArgMetCysValG 440
      |          ::: ||||| ||||| ||||| ||||| |||||
1267 CCGTTCATATGTGTTCACACCGCTTTGGTGGAGGCCAAGGCTATGTCGG 1316
      || ||||| ||||| ||||| ||||| ||||| ||||| |||||
440 lYlYSgluPhAlaLysValLeuLeuLysIlePheLeuValGluLeuThr 456
      || ||||| ||||| ||||| ||||| ||||| ||||| |||||
1317 GTTACGAGCGTGGCTAGGTTGCATCTCTGTTTCCTTCACCGCTAGTG 1366
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
457 GlnHisCysAsnTrpIleLeuSerAsnGlyProProthrmetyThrGl 473
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
1367 ACAGCGTTCACCTTGCGTT.....CCTGCAGACGACAA 1401
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
473 yProthrlleTyProValAspAsnLeuProThrLysPheThrSerTyrv 490
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
1402 GCTGGTTTCTTCCACTACACAGACGACAAACGCTACCGCATCTTCG 1451
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
490 a)Arg 491
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
1452 TGAGG 1456

seq_name: /sids1/gcgdata/geneseq/geneseqn/MA2000.DAT.AAC47986
seq_documentation_block:
ID AAC47986 standard; DNA; 1649 BP.
AC AAC47986;
AC 18-OCT-2000 (first entry)
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 55845.
XX
XX Hybridisation assay; genetic mapping; gene expression control;
XX protein identification; signal transduction pathway;
XX metabolic pathway; promoter; termination sequence; ss.
OS Arabidopsis thaliana.
XX
XX EPI033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 19-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
XX 23-APR-1999; 99US-0130891.
XX 28-APR-1999; 99US-0131449.
XX 30-APR-1999; 99US-0132048.
XX 30-APR-1999; 99US-0132407.
XX 04-MAY-1999; 99US-0132484.
XX 05-MAY-1999; 99US-0132485.
XX 06-MAY-1999; 99US-0132486.
XX 06-MAY-1999; 99US-0132487.
XX 07-MAY-1999; 99US-0132863.
XX 11-MAY-1999; 99US-0134218.
XX 14-MAY-1999; 99US-0134219.
XX 14-MAY-1999; 99US-0134221.
XX 14-MAY-1999; 99US-0134370.
XX 18-MAY-1999; 99US-0134768.
XX 19-MAY-1999; 99US-0134941.
XX 20-MAY-1999; 99US-0135124.
```

```
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 08-JUN-1999; 99US-0137724.
PR 10-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0138847.
PR 16-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 08-JUL-1999; 99US-0142390.
PR 09-JUL-1999; 99US-0142803.
PR 12-JUL-1999; 99US-0142920.
PR 13-JUL-1999; 99US-0142977.
PR 14-JUL-1999; 99US-0143542.
PR 15-JUL-1999; 99US-0143624.
PR 16-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 19-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145221.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
```

```
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160981.
PR 25-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
```

```
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

alignment_scores:
    quality: 552.50      length: 485
    ratio: 1.811        gaps: 14
Percent Similarity: 62.887   Percent Identity: 29.897

alignment_block:
US-09-668-482-2 x AAC47986 ..

Align seg 1/1 to: AAC47986 from: 1 to: 1649

19 LeuLeuPheLeuAlaIaValIysLeuTrpGluMetLeuMetIleArgR 35
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
79 CTCCTCCTCCTCTTCATCCATCCGCCGCCGCTTCCTCTACTCCGCCG 128
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
129 TACACGTTACCGT...CGATGGGTCTCTCCGCCGAGCGCTGCTCC 175
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
52 ropheIleGIyGIuThrLeuGlnLeuIle.....LeuGln 63
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
176 CTCTGATAGAGAGACTTTCACGTCGATCGAGCTTACAAACAGAGAAC 225
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
64 ArgArgIysPheLeuArgMetIysArgGlnIysTyrGIyCysIleTyr 80
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
226 CCTGAGCCTTTCATCGACGAGAGAGTAGCCCGGATACGGTTGGTTTCA 275
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
80 sThrHisLeuPheGlyAsnProThrValArgValMetGlyAlaAspAsn 97
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
276 GACGCATCTTTTGGTGACGAGCATTTCTCAGTCGACCCGGAACGA 325
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
97 aIArgGlnIleLeuGlyGlnHisIysLeuValSerValGlnTrpPro 113
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
326 ACCGTTTGTCTTCAGACGAGGAAAGCTTTTGAAGTCTTATCCP 375
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
114 AlaSerValArgThrIleLeuGlySerAspThrLeuSerAsnValHisG 130
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
376 GCTTCACATTTTGAACCTTTTGGGAAACACTCTGCTCTTATGAAGG 425
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
130 yValGlnHisIysAsnIysIysIysAlaIleMetArgAlaPheSerArg 147
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
426 TTCCTTGCATAAA...CGTAGCACCTCTCACCATCATGCTTGCTAAT 472
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
147 spAlaLeu.....GluHisTyrIleProValIleGlnGlnIuValIys 161
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
473 CTTCATATCATTAAGACCATCTCATGCTTGAATTAACCGGTTAGTCGG 522
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
162 SerAlaIleGlnIuTrpLeuGlnIysAspSerCysValLeuValTyrPr 178
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
523 TTTATCTTGATTTCTGG.....TCCCTCGTGTTCCCTCAGAGA 563
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
178 oGluMetIysIysLeuMetPheArgIleAlaMetArgIleLeuLeuGIy 195
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
564 AGAAGCAAAAAAGATTAACGTTTGAGCTTAACGTAAGCAGTTGAGCT 613
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
195 heGluPro.....GluGlnIleIysThrAspGluGlnIuLeu 207
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
614 TTGATCCAGGGGAAATGAGAGTGAGCTTTAAGGAAAGATATCTTCTG 663
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
208 ValGluAlaPheGluGluMetIleIysAsnLeuPheSerIleuProIleAs 224
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
664 ATCGAAGGCTTC.....TTCCTCTTCCCTCC 692
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
224 yValProPheSerGIyLeuTyrArg...GlyLeuArgAlaArgAsnPheI 240
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
693 TCTC...TTCTCCACCCTTACCGCAAAAGCCATCCAAAGCGGAGGAGG 739
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
240 leHisSerIysIleGluGlnAsnIleArgIysIleGlnAspAsp 256
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
```



```
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145112.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145224.
PR 27-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145513.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
```

```
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0156458.
PR 28-SEP-1999; 99US-0156459.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159325.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159638.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 25-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
```

alignment_scores:

Quality:	481.50	Length:	484
Ratio:	1.649	Gaps:	13
Percent Similarity:	60.331	Percent Identity:	26.860

alignment_block:

US-09-668-482-2 x AAC48714 ..

Align seg 1/1 to: AAC48714 from: 1 to: 1398

```
1 MetGlyLeuTyrrThrLeuMetValThrPheLeuCystrIleValLeuPr 17
|||||
22 ATGGGTCTTCCTTGATCATCGTCT..CTCTGTTCCCTCTCCT.. 66
17 ovalLeuLeuPheLeuAlaIaValLysLeuTrpGluMetLeuMetIleA 34
|||||
67 .....CGATGGAATCAGATCGAAT. 87
34 rGaTgValAspProAsnCysArgSerProLeuProPogLyThrMetGly 50
88 .....ACCAAGATGCTCTCTCTCGAACCAGGCG 120
51 LeuProPheIleGlyIuThrLeuGlnLeuIleLeuGlnArgArglySph 67
|||||
121 TGCCCAATCTTTGGCGAACAACCGAGATTCTCAACAAGGCCCAACT 170
```

```

1020 AFTGGCAACGATCGTTAATGGGCTCTTAAGCAAAACTACTCGTGAATTGG 1069
378 luleuAsnGlyTTCGlnIleProLysGlyTPAsnValIleTyrSerile 394
|||||
1070 AAATCAACAGCGTTATTAAATCCCAAAAGATGAGAAATTATATGAACACG 1119
395 CysAspThrHisAspValAlaAspValIleProAsnLysGluGluPheG1 411
|||||
1120 AGGCAATTAATTATACGATGCATAATCTTATATGACAGCCATTGATCTTTAA 1169
411 nProGluArgPheMetSerLysGlyLeuGluAspLysSerArgPheAsnT 428
|||||
1170 TCCATGCGATGGATGAGAAAGACCTTGAG.....TCACAAACATCAT 1213
428 TyrIleProPheGlyGlyLysSerArgMetCysValGlyLysGluPheAla 444
|||||
1214 GCTTGTGCTTTGGAGGAGTGGACAAAGCGCTTGTCTCGTGAAGCAACTAGCG 1263
445 LysValLeuLeuLysIlePheLeuValGluLeuThrGlnHisCysAsnTr 461
|||||
1264 ATTGTCCGATCTCGAGCGTCTCTCCATTACTTGTGTACGAGTATACAGATG 1313
461 PileuSerAsnGlyProProThrMetLysThrGlyProThrIleTyrP 478
|||||
1314 GGAGGAATATAGAGAGGCGATGAATTAATG.....GTCCTTTC 1348
478 ro 478
||
1349 CG 1350

seq_name: /SIDSI/gcgcdata/geneseq/geneseqn/NA2000.DAT: AAC35582
seq_documentation_block:
ID AAC35582 strand: DNA; 1667 BP.
XX
XX AAC35582;
XX
XX 17-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 10705.
XX
XX Hybridisation assay; genetic mapping; gene expression control;
XX protein identification; signal transduction pathway;
XX metabolic pathway; promoter; termination sequence; ss.
XX Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
XX PN
XX 06-SEP-2000.
XX
XX PF 25-FEB-2000; 2000EP-0301439.
XX
XX XX 25-FEB-1999; 99US-0121825.
XX PR 05-MAR-1999; 99US-0123180.
XX PR 09-MAR-1999; 99US-0123548.
XX PR 23-MAR-1999; 99US-0125788.
XX PR 25-MAR-1999; 99US-0126264.
XX PR 29-MAR-1999; 99US-0126785.
XX PR 01-APR-1999; 99US-0127462.
XX PR 06-APR-1999; 99US-0128234.
XX PR 08-APR-1999; 99US-0128714.
XX PR 16-APR-1999; 99US-0129845.
XX PR 19-APR-1999; 99US-0130077.
XX PR 21-APR-1999; 99US-0130449.
XX PR 23-APR-1999; 99US-0130510.
XX PR 23-APR-1999; 99US-0131449.
XX PR 28-APR-1999; 99US-0132048.
XX PR 30-APR-1999; 99US-0132407.
XX PR 04-MAY-1999; 99US-0132484.
XX PR 05-MAY-1999; 99US-0132485.

```


PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143562.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.

PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151086.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.

```

PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161820.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

alignment_scores:

Quality:	475.50	Length:	483
Ratio:	1.628	Gaps:	15
Percent Similarity:	60.455	Percent Identity:	27.122

alignment_block:

US-09-668-482-2 x AAC35582

Align seg 1/1 to: AAC35582 from: 1 to: 1667

```

1 MetGlyLeuTyrThrLeuMetValThrPheLeuGlyThrIleValLeuPr 17
   ::::::::::::::::::::::::::::::::::::::::::::::::::::
109 TTGGCTCTCTCTGTCATCATTTGTTGT...TTATGTCCTGCTCTTCTC.. 153
17 ovalLeuPheLeuAlaValLysLeuTTPGLeuMetLeuMetIleA 34
   ::::::::::::::::::::::::::::::::::::::::::::::::::::
154 .....CGATGGACACAGATGCATAT. 174
34 rGArGValAspProAsnGysArgSerProLeuProProGlyThrMetGly 50
   ::::::::::::::::::::::::::::::::::::::::::::::::::::
175 .....TCTAGAAAGCTCTTCTCTCTGAGACATGGGC 207
51 LeuProPheIleGlyIuThrLeuGlnLeuIleLeuGlnArgArgLysPh 67
   ::::::::::::::::::::::::::::::::::::::::::::::::::::
208 TGCCCATATATTGGTGAACACACAGATTCTTAAACAAGACAGATTT 257
67 eleuArgMetLysArgGlnLysTyrGlyCysIleTyrLysThrHisLeuP 84
   ::::::::::::::::::::::::::::::::::::::::::::::::::::
258 CATGAAAAACCAAGACTAAGATATGGAGCTTCTTCAAGCTTCACATTC 307
84 heGlyAsnProThrValArgValMetGlyAlaAspAsnValArgGlnIle 100
   ::::::::::::::::::::::::::::::::::::::::::::::::::::
308 TTGGCTGCCCAACAATAGTCTCAATGACGACAGATTAAACAGATTCATT 357
101 LeuLeuGlyIuHisLysLeuValSerValGlnTTPProAlaSerValAr 117
   ::::::::::::::::::::::::::::::::::::::::::::::::::::
358 CTATATGAATGAATCGAAAGAGCTTGTGCCGTATACCCGCAATATAGCT 407
117 gThrIleLeuGlySerAspThrLeuSerAsnValHisGlyValGlnHisL 134
   ::::::::::::::::::::::::::::::::::::::::::::::::::::
408 TGTATTTCTTAGGACATGCACATAGCTGCGGTTCAATGCGCCAGCCACC 457
134 ysAsnLysLysLysAlaIleMetArgAlaPheSerArgAspAlaLeu... 149
   ::::::::::::::::::::::::::::::::::::::::::::::::::::
458 GCGTATATGAGAGCGCTGCTGCTTTCTTATATAGCCCAACCATGATGAA 507
150 GlnHisTyrIleProValIleGlnGlnIuValLysSerAlaIleGlnI 166
   ::::::::::::::::::::::::::::::::::::::::::::::::::::
508 GACCATCTCTTGCTTAAGATTGATGATTCAAGAAATATCTTTGCGG 557
166 uTPLeuGlnLysAspSerCysValLeuValTyrProGlnMetLysLysL 183
   ::::::::::::::::::::::::::::::::::::::::::::::::::::
558 TTGGGATGATCTTGAGACA...GTTGATATCCAAAGAAAGACCAACATA 604
183 euMetPheArgIleAlaMetArgIleLeuLeuGlyPheGluProGlnIu 199
   ::::::::::::::::::::::::::::::::::::::::::::::::::::
605 TGGCATTT.....TTATCATCCTTGTTCACAAATAGCTGAGACTTGG 645

```

```

200 IleLysThrAspGluGlnGluLeuValGluAlaPheGluGluMetIleL 216
   ::::::::::::::::::::::::::::::::::::::::::::::::::::
646 AAAAAACCAAGAGGTTGAAGAATATAGAACACAGATTTTCAAGCTTGT 695
216 sAsnLeuPheSerLeuProIleAspValProPheSerGlyLeuTyrArg 233
   ::::::::::::::::::::::::::::::::::::::::::::::::::::
606 GGAACCTCTATCCGTCGCCGATGATATCCCGGACAGAAATTAACCCAG 745
233 lLeuArgAlaArgAsnPheIleHisSerLysIleGluIuAsnIleArg 249
   ::::::::::::::::::::::::::::::::::::::::::::::::::::
746 GAGTCAAGCAAGAAACAACATC.....GATAGCTTATTTG 780
250 LysLysIleGlnAspAspAsnGlnLysGlnIuLysTyrLysAspAl 266
   ::::::::::::::::::::::::::::::::::::::::::::::::::::
781 ACAGACTGATGCAGAAAGAGTCTGGAGAACTTCACAGACAT 830
266 aleuGlnLeuLeuIleGlnAsnSerArgArgSerAspLysProPheSer 283
   ::::::::::::::::::::::::::::::::::::::::::::::::::::
831 GTTGGGTACTTGATG.....AAGAGAGATTAACCGATTAATTGCT 871
283 eu.....GlnAlaMetLysGluAlaIleThrGluLeuLeuPheGly 297
   ::::::::::::::::::::::::::::::::::::::::::::::::::::
872 TAACGATTAAGAGATAGAGATCAAGTGAAGATCTGATATTCGCGT 921
298 HisGlnThrThrAlaSerThrAlaThrSerLeuValMetPheLeuGly 314
   ::::::::::::::::::::::::::::::::::::::::::::::::::::
922 TATAGACTGTCTCTACACCTCCATGATGCTCTTAAATCTCCATGA 971
314 uAsnThrGluValAlaGlnLysValArgGluIu.....ValGln 328
   ::::::::::::::::::::::::::::::::::::::::::::::::::::
972 TCATCCAAAGCTCTTGAGAACTCAGAGAAACATTGGCTTAAGGG 1021
328 lLysValGluMetGlyMetLysTyrProGlyLysGlyLeuSerMetGlu 344
   ::::::::::::::::::::::::::::::::::::::::::::::::::::
1022 AGAGAAAAACGA.....CCTGACGAACCGCTCCTCCTGAC 1056
345 LeuLeuAspGlnLeuLysTyrThrGlyCysValIleLysGluThrLeu 361
   ::::::::::::::::::::::::::::::::::::::::::::::::::::
1057 GATTTTAATCGATGAATTCACCTCGAGCTGTGATCTTTGAGACATCAAG 1106
361 gIleAsnProProValProGlyLysPheArgValAlaLeuLysThrPhe 378
   ::::::::::::::::::::::::::::::::::::::::::::::::::::
1107 ATTGGCAACGATGTATTATGTGCTCTAGGAAACATACACACACTTAA 1156
378 lLeuAsnGlyTyrGlnIleProLysGlyTyrAsnVal...IleTyrSer 393
   ::::::::::::::::::::::::::::::::::::::::::::::::::::
1157 AACTCAACGGTTATTATATCCCAAAAGGTTGGAGAAATTTACGATTAACA 1206
1207 AGAGAGATTATACATATGATACATC...CTTTATGAGATATCAATGATCT 1253
410 eGlnProGlnArgPheMetSerLysGlyLeuGluAspGlySerArgPheA 427
   ::::::::::::::::::::::::::::::::::::::::::::::::::::
1254 TAACCATGAGATGATGAGAAAGAGCTTAAGATCAAAAGCTATATTTC. 1302
427 snTyrIleProPheGlyGlySerArgMetCysValGlyLysGluPhe 443
   ::::::::::::::::::::::::::::::::::::::::::::::::::::
1303 .....TTACTCTTGGAGGTGGAGTTAGGCTTGGCCCTGGAAAGGAACTA 1347
444 AlaLysValLeuLeuLysIlePheLeuValGluLeuGlnHisCysAs 460
   ::::::::::::::::::::::::::::::::::::::::::::::::::::
1348 GGAATCTCGGAAGTCTCAACGCTTCTTCACTACTTTGTTCAAAATATAG 1397
460 nTPIleLeuSerAsnGlyProProThrMetLysThrGlyProThrIle 476
   ::::::::::::::::::::::::::::::::::::::::::::::::::::
1398 ATGG...GAAGAGATGAGAGAACAAATTAATGATGCTTCCAAAGGTT 1443
seq_name: /SID1/gcdata/geneseq/geneseqn/NA2000.DAT:AAAC39029
seq_documentation_block:
ID AAC39029 standard; DNA; 1685 BP.

```

XX AAC39029;
AC 17-OCT-2000 (first entry)
XX
XX
DT Arabidopsis thaliana DNA fragment SEQ ID NO: 23101.
XX
DE Hybridisation assay; genetic mapping; gene expression control;
XX protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EPI03405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.

PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.

```
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151330.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

alignment_scores:
Quality: 470.00 Length: 483
Ratio: 1.632 Gaps: 9
Percent Similarity: 59.627 Percent Identity: 26.915
```

alignment_block:
US-09-668-482-2 x AAC39029 ..

Align seg 1/1 to: AAC39029 from: 1 to: 1685

```
14 llevalleuprovallleuleupheleualalaValLysLeuTrpDlume 30
:::|||||:::|||||:::
139 TTGGTTTGACCAATCCATCTTATGCTTG.....CTTTGCTAG 179
30 lleuMetleargValaspproasnCysargserProleupProProG 47
:::|||||:::|||||:::
180 AGTGATTGTTTCAAGAAAGAAAGACAGTAGAGTAACTTCTCTCG 229
47 lyThMetGlyLeuProheleleGlyluThrLeuGlnleuTleuGln 63
```

```
|||||:::|||||:::|||||:::|||||:::|||||:::
230 GTCCATGAGATGGCCTTACTTAGAGAGACTCTCAACTATTCACAA 279
64 Arg...ArgLysPheLeuArgMetLysArgGlnLysYrGlyCysIleTty 79
:::|||||:::|||||:::|||||:::|||||:::
280 AACCCCAATGTTTCTTCACTCCCAAGCAAAAGAGATATGAGACATAT 329
79 llySThrHisLeuPheGlyAsnProThrValArgValMetGlyAlaAsp 96
:::|||||:::|||||:::|||||:::|||||:::
330 CAAACCCCAATCCCGGCTATCCATCGCGTATGTTGGCTAGCCCTGAG 379
96 snValArgGlnlleuLeuGlyGlnHisLysLeuValSerValGlnTrp 112
CTCCGAGCTTTGTTACTTGTGACTATGCCATATGTTCAACCACTTAT 429
113 ProAlaSerValArgThrIleLeuGlySerAspThrIleSerAsnVal 129
||| |||:::|||||:::|||||:::|||||:::
430 CCGAGAAGCAAAAGAGAGCTGATAGACCCCTGCACCTTTTCCACCA 479
129 sGlyValGlnHisLysAsnLysLysLysAlaIleMetArgAlaPheSera 146
:::|||||:::|||||:::|||||:::|||||:::
480 AGGAGATTATCATCTCCATATAGGAAACTGTCAATCCCTTCTTACC 529
146 rGAspAlaLeuGlnHisTyrIleProValIleGlnGlnGluValLysSer 162
:::|||||:::|||||:::|||||:::|||||:::
530 CTGAACCATCCGTAACATCATCCCTGATATCGACACATTCGCCCTTCT 579
163 AlaIleGlnGluTrpLeuGlnLysAspSerCysValLeuValYrProG 179
:::|||||:::|||||:::|||||:::|||||:::
580 TCCTTACATCTTGGGCCAATATG...CCGATTGCTCCACCTTCCACGA 626
179 uMetLysLysLeuMetPheArgIleAlaMetArgIleLeuLeuGlyPheG 196
|||||:::|||||:::|||||:::|||||:::
627 GATGAGAGATTCGCCCTTGTATGGGTATTCAGCCATTTTGA... 672
196 luproGlnGlnIleLysThrAspGlnGluLeuValGlnAlaPheGln 212
673 .....CATTTGGAGACTTCTTCAAA 693
213 GluMetIleLysAsnLeuPhe.....SerLe 221
694 GAGATTCGTGAACATACATCAATATTTGGCAAAAGCTTCAACACTT 743
221 uproIleAspValProPheSerGlyLeuYrArgGlyLeuArgAlaArg 288
744 CCCCATGAGTCTCCCGGAACATCTTATCAAAAGCTTCATGCGCAGAA 793
238 snPheIleHisSerLysIleGlnGlnAsnIleArgLysLysIleGlnAsp 254
:::|||||:::|||||:::|||||:::|||||:::
794 AGCAGCTAAAGACGATAGTAAGCAGATTAATATGCGAAAGA..... 834
255 AspAspAsnGlnAsnGlnLysTyrLysAspAlaLeuGlnleuLeuI 271
835 .....AGAGAGAAAAGGCCCTTGCAACGACTTCTTGGTCATCTACT 878
271 eGluAsnSerArgArgSerAspGluProPheSerLeuGlnAlaMetLysG 288
:::|||||:::|||||:::|||||:::|||||:::
879 CAACCTTCAGAAAGCAAAAGGTCGTGCTTAACCCAGAAAGAGATTGCG 928
288 lualAlaThrIleuLeuPheGlyGlyHisGluThrAlaSerThr 304
:::|||||:::|||||:::|||||:::|||||:::
929 ACACATCATCGAGTCTTTCGCGACAGACACGACACTGATTCG 978
305 AlaThrSerLeuValMetPheLeuGlyLeuAsnThrGlnValValGln 321
979 TTAACCTTGATTCCTTAAGTACTTACATGATGATGCAAACTTCTGAAC 1028
321 sValArgGlnGluValGlnGlnLysValGlnMetGlyMetYrThrProG 338
|||||:::|||||:::|||||:::
1029 TGTTAAGGCTGAG...CAAAAGCTATATATGAGAAACAACTAGAGGA 1075
338 lYlThMetGlyLeuSerMetGluLeuLeuAspGlnLeuLysTyrThrGlyCys 354
||| |||:::|||||:::|||||:::|||||:::

```

```
1076 AGAAACCTTTAACTGAGACAAACGAGAAATATGCCACTGACACATAG 1125
355 ValIleuysglutThrLeuArgIleasnProValProGlyGlyPheAr 371
||||| ||||| ||||| |||||
1126 GTTATAGTTGAAGCTGAGAGATGCAAGCATCATATCCTTCACATTCAG 1175
371 gValAlaLeuLysThrPheGluLeuAsnGlyTyrGlnIleProLysGlyT 388
| ||||| ||||| ||||| ||||| |||||
1176 AGAAGCAGTGGTTCATCTGAAATATAGGATATTTGATACCTAAGGAT 1225
388 rPasnValIleTyrSerIleCysaspThrHisAspValAlaAspValPhe 404
||||| ||||| ||||| |||||
1226 GGAAGTGAATGCCACTGTTGCGAATATTCATCACAATCCGAAATATT 1275
405 ProAsnLysGluGluPheGlnProGluArgPheMetSerLysGlyLeuG 421
||| ||| ||||| ||||| |||||
1276 TCAAAACCTGTAGGTTTCGACCCATCTAGATTC.....GA 1310
421 uAspGlySerArgPheAsn...TyrIleProPheGlyGlySerArgm 437
| : : : : ||| : : : : ||||| : : : :
1311 GGTAAATCCGAAAGCCGAATACATTCATGCCCTTTGGAAGTGGAGTTCATG 1360
437 etCysValGlyLysGluPheAlaLysValLeuLeuLysIlePheLeuVal 453
||| ||| ||||| ||||| |||||
1361 CTTGTCGCCGGAACGAACTGCCAAGTTACAATCTTATATTTCTCCAC 1410
454 GluLeuThrGlnHisCysAsnTrpIleLeuSerAsnGlyProProThrMe 470
: : : : : : : : : : : : : : : : : :
1411 CATTAAGTTCCAAATTCGATGCGAAGTGAAGGAGAGAGAAAGGAAT 1460
470 tLysThrGlyProThrIleTyrProValAspAsnLeuProThrLysPhe 486
: : : : : : : : : : ||| : : : : : : : :
1461 ACAGTACAGTCCATTCCAAATACCTCAAAACGGTCTCCCGCTACATTT 1509
```



```

238 TTGCGCTTCATTGGAGAAACGCTCCAGCTGATCTCCAGAGAAGAAAGTT 287
67  eLeuArgMetLysArgGlnLysTyrGlyCysIlePheTyrThrHisLeuP 84
288 TCTGGCGCATGAAACGGCAGAAATACGGGTGCTACACAGACGCACTCT 337
84  heGlyAsnProThrValArgValMetGlyAlaAspAsnValArgGlnIle 100
338 TCGGGAACCCGCACTGTCAAGGGTGATGGAGCGCATATGTGAGCGAGATT 387
101 LeuLeuGlnGlnHisLysLeuValSerValGlnTrpProAlaSerValAr 117
388 CTGCTGGGCGGACACACACAGCTGTCTCTTCAGTGGCCAGCATCAGTGAG 437
117 gThrIleLeuGlnLysSerAspThrLeuSerAsnValHisGlyValGlnHis 134
438 AACCATCTGGGCTGTGACACCTCTCCAAATGTCATGTGAGTCAACACA 487
134 ysaAsnLysLysLysAlaIleMetArgAlaPheSerArgAspAlaLeuGln 150
488 AAAACAAGAAAAAGCCCATTTATGAGGGCGTCTCTCGAGATGCTCTGGAG 537
151 HisTyrIleProValIleGlnGlnGlnValLysSerAlaIleGlnGlnTr 167
538 CACTACATCCCGTGATCCAGCAGAGAGGTGAAGCCGCAATACAGAAATG 587
167 pLeuGlnLysAspSerCysValIleuValTyrProGlnuMetLysLysLeu 184
588 GCTCAAAAAGACCTCGCGCTGCGTGTATCCAGAAATGAAAGAACCACTCA 637
184 etPheArgIleAlaMetArgIleLeuLeuGlnGlyPheGluProGlnGlnIle 200
638 TGTTCGGATAGCATATGAGATCTCTGCTGGTTTGAACACAGACCAATA 687
201 LysThrAspGlnGlnGlnIleuValGlnAlaPheGlnGlnuMetLysLys 217
688 AAGCGACGACGACGAGAACTGTGAAAGCTTTGAGGAAATGATCAACAAA 737
217 nLeuPheSerLeuProIleAspValProPheSerGlyLeuTyrArgGlyL 234
738 CTGTGCTCTCTTGGCAATCGACGTTCTTTCAGTGTGCTGTACAGGGGTT 787
234 euArgAlaArgAsnPheIleHisSerLysIleGlnGlnuAsnIleArgLys 250
788 TGAAGGCGACCAATTTCATTCACCTCCAAATTTGAGAAACATCAGGAAG 837
251 LysIleGlnAspAspAsnGlnuAsnGlnuLysTyrLysAspAlaIle 267
838 AAAATTCAGATGACGACCATGAAACGAAACAGAAATACAAAGAGCCCT 887
267 uGlnLeuLeuIleGlnuAsnSerArgArgSerAspGluProPheSerLeu 284
888 TCAGCTGTTCATCGAGAACGACAGAAAGAAAGTGAACCTTTTAGTTGCG 937
284 IlaIleLysGlnuAlaIleThrGlnLeuLeuPheGlyGlyHisGlnuThr 300
938 AGCGCATGAAAGAGAGCTACAGAGCTTCTATTTGGAGGTGATGAAGACC 987
301 ThrAlaSerThrAlaThrSerLeuValMetPheLeuGlnLysuAsnTrp 317
988 ACCGCGACGACCTCAACCTCATTGTCATGTTCTGGGTGTGAACACAGA 1037
317 uValValGlnLysValArgGlnuValGlnGlnLysValGlnuMetGly 334
1038 AGTGAGTGCAAGAGGTGAGAGAGAGGTTTCAGAGAAAGTTGAATGGGCA 1087
334 eTyrThrProGlnLysGlyLeuSerMetGlnuLeuAspGlnuLys 350
1088 TGTATACACTGGAAGGCTTGAGATGAGAGCTGTTGAGCAGCTGAAG 1137
351 TyrThrGlyCysValIleLysGlnuThrLeuArgIleAsnProProValPr 367
1138 TACACTGATGTGTATTAAGAGACTTTAGAAATCAACCTCTCTGTTC 1187

```

```

367 oGlyIlePheArgValAlaLeuLysThrPheGlnLeuAsnGlyTyrGlnI 384
1188 CCGAGGATTCAGAGTCGCACTCAAAACCTTTGAATTGAATGGTTACAAA 1237
384 leProLysGlyTrpAsnValIleTyrSerIleCysAspThrHisAspVal 400
1238 TTCCTTAAGATGAGAACGTCATTTACAGCATCTGTGACACGACGATGTG 1287
401 AlaAspValPheProAsnLysGlnuPheGlnProGlnuArgPheMetSe 417
1288 GCCGACGCTCTTCCAAACAAAGAGAGATTCCAGCGGAGAGATTATGAG 1337
417 rLysGlnLeuGlnuAspGlySerArgPheAsnTyrIleProPheGlyGly 434
1338 CAAAGTCTGGAGGAGCGGTCCAGGTTTAACATCAATCCCTTCGAGGAG 1387
434 LysSerArgMetCysValGlyLysGlnuPheAlaLysValLeuLeuLysIle 450
1388 GATCCAGATGTGTGTGGGCAAAAGAGTTCCCAAAAGTGTACTCAAGATC 1437
451 PheLeuValGlnLeuThrGlnHisCysAsnTrpIleLeuSerAsnGlyPr 467
1438 TTTTAGTTGATTAACGACGATTCGAATTGGATTCTCTCAAAAGGAGCC 1487
467 oProThrMetLysThrGlyProThrIleTyrProValAspAsnLeuProT 484
1488 CCCGACAAATGAAACAGGCGCGACTATTATTACCGATGAGCAATCTCCCTA 1537
484 hTrLysPheThrSerTyrValArgAsn 492
1538 CCAATTCCTAGTTATGTCAAGAAAT 1563

seq_name: /cgn2.6/prodata/2/ina/09.COMB.seq:us-08-882-164D-3
seq_documentation_block:
; Sequence 3, Application us/08882164D
; Patent No. 6306624
; GENERAL INFORMATION:
; APPLICANT: Petkovich, P. Martin, White, Jay A.,
; APPLICANT: Beckett, Barbara R., Jones, Glenville
; TITLE OF INVENTION: Retinoid Metabolizing Protein
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Blake, Cassels & Graydon
; STREET: Box 25, Commerce Court West
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5L 1A9
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 MB storage
; COMPUTER: COMPAQ, IBM PC compatible
; OPERATING SYSTEM: MS-DOS 5.1
; SOFTWARE: WORD PERFECT
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/882,164D
; FILING DATE: June 25, 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/667,546
; FILING DATE: June 21, 1996
; APPLICATION NUMBER: 08/724,466
; FILING DATE: October 1, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Hunt, John C.
; REGISTRATION NUMBER: 36,424
; REFERENCE/DOCKET NUMBER: 50767/00010
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 863-4344
; TELEFAX: (416) 863-2653
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1850 base pairs

```



```

;       TYPE: nucleic acid
;       STRANDEDNESS: single
;       TOPOLOGY: linear
;       US-08-882-164D-3

```

```

alignment_scores:
    Quality: 2563.00      Length: 492
    Ratio: 5.209          Gaps: 0
    Percent Similarity: 100.000    Percent Identity: 100.000

```

```
alignment_block:
```

```
US-09-668-482-2 x US-08-882-164D-3 ..
```

```
Align seg 1/1 to: US-08-882-164D-3 from: 1 to: 1850
```

```

1 MetGlyLeuTYrThrLeuMetValThrPheLeuGlyThrIleValLeuPr 17
  |||||
88 ATGGGGCTGTACACCCCTTATGTGCTACCTTCTGACCATCGTCTAC 137
  |||||
17 ovalLeuLeuPheLeuAlaIleValLysLeuTrpGluMetLeuMetIle 34
  |||||
138 CGTTTACTCTTCTCGCGCGGTGAAGTGTGGAGATGTTAATGATCC 187
  |||||
34 rGArValAspProAsnGlySerGlySerProLeuProProGlyThrMetGly 50
  |||||
188 GACGAGTGCATCCGAACTCAGAAAGTCTTACCGCCAGGTACCATGGGC 237
  |||||
51 LeuProPheIleGlyGluThrLeuGlnLeuIleLeuGlnArgArgLysPh 67
  |||||
238 TTGCGCTTATGTGAGAAACGCTCCAGCTGATCTCCAGAGAGAAAGTT 287
  |||||
67 eLeuArgMetLysArgGlnLysTyrGlyCysIleTyrLysThrHisLeuP 84
  |||||
288 TCTGGCCATGAAACGGCAGAAATACGGGTGCATCTACAAAGACGACCTCT 337
  |||||
84 heGlyAsnProThrValArgValIleMetGlyAlaAspAsnValArgGlnIle 100
  |||||
338 TCGGGAAACCGACTGTACAGGTGATGGAGCTGATATGTGAGGAGATT 387
  |||||
101 LeuLeuGlyGluHisLysLeuValSerValGlnTrpProAlaSerValAr 117
  |||||
388 CTGCTGGCGGAACACAGCTGGTGTCTGTTCAGTGGCCAGCATCAGTAG 437
  |||||
117 gThrIleLeuGlySerAspThrLeuSerAsnValHisGlyValGlnHisL 134
  |||||
438 AACCATCTGGGCTGTGACACCTCTCCATGTCCATGTGAGATTCAACACA 487
  |||||
134 yAsnLysLysLysAlaIleMetArgAlaPheSerArgAspAlaLeuGln 150
  |||||
488 AAAACAGAAAAGGCCATTATGAGGGCGTCTCTGAGATGCTCTGGAG 537
  |||||
151 HisTyrIleProValIleGlnGlnGluValLysSerAlaIleGlnGlnL 167
  |||||
538 CACTACATTCCTCGTATCCAGCAGAGGTGAAGAGCGCCATACAGGAATG 587
  |||||
167 pLeuGlnLysAspSerCysValLeuValTyrProGluMetLysLysLeuM 184
  |||||
588 GCTGCAGAAAAGACTCCTGCGTGTGTTATCCAGAAATGAAGAAACTCA 637
  |||||
184 ePheArgIleAlaMetArgIleLeuLeuGlyPheGluProGlnGlnIle 200
  |||||
638 TGTTCGCGATAGCTATGAGAAATCTGCTGTGTTTGACACAGGCAAAAT 687
  |||||
201 LysThrAspGluGlnGluLeuValGluAlaPheGluGluMetIleLysAs 217
  |||||
688 AAGACGAGCAGCAAGAACTGTGGAAGCTTTGAGGAAATGATCAAAA 737
  |||||
217 nLeuPheSerLeuProIleAspValProPheSerGlyLeuTyrArgGlyL 234
  |||||
738 CTGTGTTCTCTGCAATGACAGCTTCTTCAGTGGTCTTACAGGGGTT 787
  |||||
234 euArgAlaArgAsnPheIleHisSerLysIleGlnGluAsnIleArgLys 250

```

```

|||||
788 TGAAGGACGCAATTCATTCACCTCCAAAATTGAGGAAAACATCAGAGAG 837
|||||
251 LysIleGlnAspAspAsnGlnAsnGlnLysTyrLysAspAlaIle 267
|||||
838 AAATTCATCATGATGACGCAATGAAACGAAACAGAAATACAAAGACGCCCT 887
|||||
267 uGlnLeuLeuIleGluAsnSerArgArgSerAspGluProPheSerLeuG 284
|||||
888 TCACCTTGTATCGAGACACGCAAGAAAGTGAACCTTTAGTTTGG 937
|||||
284 IlnAlaMetLysGluAlaAlaThrGluLeuLeuPheGlyGlyHisGluThr 300
|||||
938 AGCGATGAAAGAAAGACGTACAGAGCTTCTATTGGAGGTGATGAAGAC 987
|||||
301 ThrAlaSerThrAlaThrSerLeuValMetPheLeuGlyLeuAsnThrG 317
|||||
988 ACCGCCAGCACTGCACACCTGATGTCATGTTCTGGGCTTGAAACACGA 1037
|||||
317 uValValGlnLysValArgGluGluValGlnGluLysValGluMetGlyM 334
|||||
1038 AGTGTGACAGAGGTTCAGAGAGAGGTTCAGAGAGAGTTGAATGGCA 1087
|||||
334 eTyrThrProGlyLysGlyLeuSerMetGluLeuLeuAspGlnLeuLys 350
|||||
1088 TGTATACACCTGGAAGGGCTTGAGTATGAGCTGTGGACACAGCTGAG 1137
|||||
351 TyrThrGlyCysValIleLysGluThrLeuArgLysAsnProProValPr 367
|||||
1138 TACACTGATGTGTGATTAAAGACCTTTAAGATCAACCTCTGTTCC 1187
|||||
367 oGlyGlyPheArgValAlaLeuLysThrPheGluLeuAsnGlyTyrGlnI 384
|||||
1188 CGGAGATTCAGAGTCCGACTCAAAACCTTTGATGATGATGTTACCAAA 1237
|||||
384 IeProLysGlyTyrAsnValIleTyrSerIleCysAspThrHisAspVal 400
|||||
1238 TTCTTAAAGATGGAACGTCATTACAGCATCTGTGACACGACGAGATGTG 1287
|||||
401 AlaAspValPheProAsnLysGluGlnPheGlnProGluArgPheMetSe 417
|||||
1288 GCCGAGCTTCTTCCAAACAAAGAGAGTTCAGCGGAGAGATTCATGAG 1337
|||||
417 rLysGlyLeuGlnAspGlySerArgPheAsnTyrIleProPheGlyGlyG 434
|||||
1338 CAAGGTCTGAGAGAGCGGCTCAGAGTTAACTACATCCCTTCGAGAGAG 1387
|||||
434 LysSerArgMetCysValGlyLysGluPheAlaLysValLeuLeuLysIle 450
|||||
1388 GATCCAGAGATGTGTGGCAAAAGGTTCCGCCAAAGTGTACTCAAGATC 1437
|||||
451 PheLeuValGluLeuThrGlnHisCysAsnTrpIleLeuSerAsnGlyPr 467
|||||
1438 TTTTATGTTGATTAACGAGCATTTGCAATTGATGATCTCTCAAAAGGACC 1487
|||||
467 oProThrMetLysThrGlyProThrIleTyrProValAspAsnLeuProT 484
|||||
1488 CCCGCAATGAAACAGAGGCCGCACTATTACCCAGTGAACAATCTCCCTA 1537
|||||
484 hLysPheThrSerTyrValArgAsn 492
|||||
1538 CCAAAATTCACGTATGTACAGAAAT 1563

```

```
seq_name: /cgn2_6/ptodata/2/lna/6A_COMB.seq:US-08-724-466B-5
```

```
seq_documentation_block:
```

```
; Sequence 5, Application US/08724466B
```

```
; Patent No. 6063606
```

```
; GENERAL INFORMATION:
```

```
; APPLICANT: Petkovich, P. Martin, White, Jay A.,
```

```
; APPLICANT: Beckett, Barbara R., Jones, Glenville
```

```
; TITLE OF INVENTION: Retinoid Metabolizing Protein
```

```
; NUMBER OF SEQUENCES: 30
```

```

CORRESPONDENCE ADDRESS:
ADDRESS: Blake, Cassels & Graydon
STREET: Box 25, Commerce Court West
CITY: Toronto
ZIP: M5L 1A9
COUNTRY: Canada
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 Mb storage
COMPUTER: COMPAQ, IBM PC compatible
OPERATING SYSTEM: MS-DOS 5.1
SOFTWARE: WORD PERFECT
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/724,466B
FILING DATE: October 1, 1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/667,546
FILING DATE: June 21, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Hunt, John C.
REGISTRATION NUMBER: 36,424
REFERENCE/DOCKET NUMBER: 50767/00004
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 863-4344
TELEFAX: (416) 863-2653
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1494 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-724-466B-5

alignment_scores:
Quality: 1738.00 Length: 497
Ratio: 4.080 Gaps: 5
Percent Similarity: 85.714 Percent Identity: 68.008

alignment_block:
US-09-668-482-2 x US-08-724-466B-5
Align seg 1/1 to: US-08-724-466B-5 from: 1 to: 1494

1 MetGlyLeuTyrThrLeuMetValThrPheLeuCysThrIleValLeuPr 17
||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
1 ATGGGGCTCCCGCGCTGCTGGCGAGTCTGCTGACCTTCTGTCGCGCC 50
17 ovalLeuPheLeuAlaIleValIleLeuTyrPheLeuMetLeuIle 34
||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
51 GCTGCTCTCTTCTGCTGCTGATCAAGCTCTGGACCTGTACTGCGTGA 100
34 rGArValAspProAsnCysArgSerProLeuProPheGlyThrMetGly 50
||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
101 GCGCGCGCGACGCGAGTTGTGCCCTCCCATTCGCCCGCGGAGCTATGGC 150
51 LeuProPheIleGlyIleuThrLeuGlnLeuIleLeuGlnArgArgIle 67
||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
151 TTCCCTCTTCTTGGGGAACCTTGACAGATGTACTGACGCGAGGAAGTT 200
67 eleuArGMetLysArgGlnLysTyrGlyCysIleTyrLysThrHisLeu 84
||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
201 CCGGACGATGAAGCGGAGAAATACGGCTTCATCTCAAGACGACATCTGT 250
84 heGlyAsnProThrValArgValMetGlyAlaAspAsnValArgGlnIle 100
||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
251 TCGGGCGCGCCACCGTACGGGTATGGCGCGGACAAATGCGCGGCATC 300
101 LeuLeuGlyGlnHisLysLeuValSerValGlnTyrProAlaSerValAr 117
||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
301 TTGCTGGGAGACGCGCTGTGTGTCGTCACATGCGCAGCGTGGTGGC 350
117 gThrIleLeuGlySerAspThrLeuSerAsnValHisGlyValGlnHisL 134
||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
```

```

351 CACCAATTCTGGGATCTGGCTGCTCTTAACCTGCAGCAGCTCTCGCACA 400
134 yAsnLysLysAlaIleMetArgAlaPheSerArgAspAlaLeuGlu 150
||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
401 AGCAGCGCAAGAGGATATGCGGGCTTCAGCCCGCAGGCACTCGAA 450
151 HisTyrIleProValIleGlnGlnValLysSerAlaIleGlnIle 167
||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
451 TGCATAGTCCCGGTATCACGAGGAAGTGGCAGACAGCTGGAGAGTG 500
167 pleuGlnLysAspSerCys.....ValLeuValTyrProGlu 180
||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
501 GCTG.....ACGTGGCGGACCGCGGCTCTGCTTACCCCGAGG 541
180 eLysLysLeuMetPheArgIleAlaMetArgIleLeuLeuGlyPheGlu 196
||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
542 TGAACGGCTCATGTTCGGAATCGCAGCGCATCTGCTGCGCTGCGAA 591
197 ProGlnGlnIleLysThrAsp.....GluGlnIleuValGluAl 210
||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
592 CCC...CACTGGCGGCGGCGGACTCCGAGCAGCAGCTTGTGAGAGC 638
210 aPheGlnIleuMetIleLysAsnLeuPheSerLeuProIleAspValProp 227
||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
639 CTTCGAGGAATGACCCCAATCTTCTCTGCTGCCATTCGACGTGCCCT 688
227 heSerGlyLeuTyrArgGlyLeuArgAlaArgAsnPheIleHisSerLys 243
||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
689 TCAGCGGCTGTACCGGGCATGAAGGCGCGGAACCTCATTCACGCGGC 738
244 IleGlnIleuAsnIleArgLysIleGlnAsp...AspAsnGluAs 259
||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
739 ATCGACGAAATTCGCGCAGATCTGCGGCTCGGCGGATCCGAGGC 788
259 nGluGlnLysTyrLysAspAlaLeuGlnLeuIleGlnAsnSerArg 276
||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
789 GGGCAGGCGTCAAGACGCGCTGACGCTGTGATGACGACACGCTGGG 838
276 rGSerAspGluProPheSerLeuGlnAlaMetLysGluAlaIleThrGlu 292
||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
839 AGAGGGAGCGGCTGACATGACAGCAGCACTAAAGCAATCTTCAACGAA 888
293 LeuLeuPheGlyGlnHisGluThrThrAlaSerThrIleThrSerLeu 309
||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
889 CTCCTCTTTGGAGACACGAAACACGCGCAGTGCAGCCACATCTGAT 938
309 lMetPheLeuGlyLeuAsnThrGluValGlnLysValArgGluLys 326
||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
939 CACTTACCTGGGCTTACCCACATGTTCTCCAGAAAGTGGAGAACAGC 988
326 aLingIleuLysValGlnMetGlyMetTyrThrProGlyLysGlyLeuSer 342
||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
989 TGAAGAGTAAGGTTTACTTTGCAAGACATCAAGCAACAACTGTGAC 1038
343 MetGluLeuLeuAspGlnLeuLysTyrThrGlyCysValIleLysGlu 359
||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
1039 ATGGAATTTTGGACACACTTAATATCATGCGGTGTTTAAAGGAGAC 1088
359 rLeuArgIleAsnProProValProGlyGlyPheArgValAlaLeuLys 376
||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
1089 CTTTCGACATGAATCCCGCAGTTCAGAGAGGTTTGGGTGCTGTAGAA 1138
376 hrPheGlnLeuAsnGlyTyrGlnIleProLysGlyTyrPheAsnValIle 392
||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
1139 CTTTGAATTAATGATACCAATTCACAAAGGCTGGAATGTATCTAC 1188
393 SerIleCysAspThrHisAspValAlaAspValPheProAsnLysGlu 409
||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
1189 AGTATCTGTATCTCATGATGTGGCAGAGATCTTCAACCAAGAGAGA 1238
409 uPheGlnProGlnArgPheMetSerLysGlyLeuGluAspGlySerArg 426
||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
1239 ATTAAATCCGACGATTCAGTCTCTCTCACCAGAGAGATTCAGGT 1288
```

```

426 heaenryrllleprophglyglyserargmetcysvalglysglu 442
|||||.....
1289 TCAGCTTCATTCATTGGAGAGGAGGCTGTAGAGCTGTAGAGCAAGAA 1338
443 phealalysvalleuleuylsllpheleuvalgluleuhrgrlnhiscy 459
|||||.....
1339 TTTGCAAAATCTTCTCAAAATATTACAGTGGAGCTGGCCAGGCAATG 1388
459 sasnrtrilleuseransglyproprothrmetlsthrglyprothri 476
|||||.....
1389 TGACTGGAGAGCTTCTAAATGAGGACTCTCAAGATGAAAAACGATCCACCG 1438
476 leryrprovalaspsanleuprothrlyserphrserTyr 489
|||||.....
1439 TGTATCCGTGTGACAAATCTCCCTGCAAGATTCACCCATTTTC 1479

seq_name: /cgnr2_6/prodata/2/lna/5B_COMB.seq:us-08-882-164D-5

seq_documentation_block:
; Sequence 5, Application US/08882164D
; Patent No. 6306624
; GENERAL INFORMATION:
; APPLICANT: Petkovich, P. Martin, White, Jay A.,
; APPLICANT: Beckett, Barbara R., Jones, Glenville
; TITLE OF INVENTION: Retinoid Metabolizing Protein
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Blake, Cassels & Graydon
; STREET: Box 25, Commerce Court West
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5L 1A9
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 Mb storage
; COMPUTER: COMPAQ, IBM PC compatible
; OPERATING SYSTEM: MS-DOS 5.1
; SOFTWARE: WORD PERFECT
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/882,164D
; FILING DATE: June 25, 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/667,546
; FILING DATE: June 21, 1996
; APPLICATION NUMBER: 08/724,466
; FILING DATE: October 1, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Hunt, John C.
; REGISTRATION NUMBER: 36,424
; REFERENCE/DOCKET NUMBER: 50767/00010
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 863-4344
; TELEFAX: (416) 863-2653
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1494 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-882-164D-5

alignment_scores:
Quality: 1738.00 Length: 497
Ratio: 4.080 Gaps: 5
Percent Similarity: 85.714 Percent Identity: 68.008

alignment_block:
US-09-668-482-2 x US-08-882-164D-5 ..
Align seg 1/1 to: US-08-882-164D-5 from: 1 to: 1494

```

```

1 MetGlyLeuTyrThrLeuMetValThrPheLeuCysThrIleValLeuPr 17
|||||.....
1 ATGGGGCTCCCGGCGCTGCTGGCCAGTGGCGCTGACCTTGCTGTGTGGC 50
17 oValLeuPheLeuAlaValLysLeuTyrGluMetLeuMetIleA 34
|||||.....
51 GCTGCTGCTCTTCTGCTGGTCGATCAAGCTCTGGACCTGACTGCTGTA 100
34 rGArValAspProAsnGlySerProLeuProGluThrMetGly 50
|||||.....
101 GCGGCGCGACCGCAGTGTGCTCCCTCCATGGCCCCCGGAGCTATGGGC 150
51 LeuProPheIleGlyLutThrLeuGlnLeuIleLeuGlnArGArGlySp 67
|||||.....
151 TTCCCTCTTCTTGGGAAACCTTGCAAGATGATCTGACGGAGGAAGTT 200
67 eleuArgMetLysArgGlnLysTyrGlyCysIleTyrLysThrHisLeu 84
|||||.....
201 CTGCAAGATGAAGCGCAGAAATACGCTTCACTACAAAGACGATCTGT 250
84 heGlyAsnProThrValArgValMetGlyAlaAspAsnValArgGlnIle 100
|||||.....
251 TCGGCGCGCCACCGTACGGGCTGATGGCGCGGACAACTGGCGCGCATC 300
301 TTGCTCGGAGACGACCGCGGTGGTGTGCTGCACCTGGCCACGTCGGGCG 350
117 gThrIleLeuGlySerAspThrLeuSerAsnValHisGlyValGlnHis 134
|||||.....
351 CACCATTCGTGGATCTGCTGCCCTCTTAACCTGCACGACTCTCTGCACA 400
134 ysAsnLysLysLysAlaIleMetArgAlaPheSerArgAspAlaLeuGlu 150
|||||.....
401 AGCAGCGAGAGAGGTGATTTAGCGGCTTTCAGCCGCGAGCAGCTGAA 450
151 HisTyrIleProValIleGlnGlnValLysSerAlaIleGlnGlnTyr 167
|||||.....
451 TGCTACGTCGCGGTGATCACCGAGAGTGGCGAGCCTGGAGCAGTG 500
167 pleuGlnLysAspSerCys.....ValLeuValTyrProGlu 180
|||||.....
501 GCTG.....AGCTGCGCGGAGCGCGGCTCTGCTTACCCGAGG 541
180 eLysLysLeuMetPheArgIleAlaMetArgIleLeuGlyPheGlu 196
|||||.....
542 TGAAGCGCTCATGTTCGAATCCGCACTGCCATCTACTGGGTGGAA 591
197 ProGlnGlnIleLysThrAsp.....GluGlnGluLeuValGluAl 210
|||||.....
592 CCC...CAACTGGCGGGGCGGAGCTCCGAGCAGACGCTTGGAGGC 638
210 aPheGlnGluMetIleLysAsnLeuPheSerLeuProIleAspValProp 227
|||||.....
639 CTTCGAGGAAATGACCCGCAATCTTCTGCTGCCATCCATCCAGCTGCCT 688
227 heserGlyLeuTyrArgGlyLeuArgAlaArgAsnPhelHisSerLys 243
|||||.....
669 TCAGCGGCTGTACCGGGCATGAAAGCGCGGAACCTATTCACGCGGC 738
739 ATCGAGCAAGACATTCGCGCAAGATCTGCGGCTGCGGCGCATCGAGGC 788
244 IleGlnGluAsnIleArgLysLysIleGlnAsp...AspAsnGlnLys 259
|||||.....
739 ATCGAGCAAGACATTCGCGCAAGATCTGCGGCTGCGGCGCATCGAGGC 788
259 nGluGlnLysTyrLysAspAlaLeuGlnLeuLeuIleGluAsnSerArg 276
|||||.....
789 GGGCCAGGCTGCAAAAGACGCGCTGACGCTGTGATCGACGACTGTGGG 838
276 rGSerAspGluProPheSerLeuGlnAlaMetLysGluAlaAlaThrGlu 292
|||||.....
839 AGAGGGAGAGCGGCTGAGCATGCGAGCAGCACTAAAGCAATCTTCAACGAA 888
293 LeuLeuPheGlyGlnIleGluThrThrAlaSerThrAlaThrSerLeuVal 309

```

```
|||||
889 CTCCTTTGGAGACAGAAACACAGGCACTGCAGCCACTCTCTGAT 938
309 lmecheleuengllyleuasnthrGluValAlnLysValArgGluIuv 326
939 CACTTACCTGGGGCTTACCCACATTTCTCCAGAAAGTGCAGAAAGAC 988
326 aAlnGluLysValAlnMetGlyMetYThrProGlyLysGlyLeuSer 342
989 TGAAGAGTAAGGCTTACTTTGCAAGACCAATCAAGACAAAGTTGGAC 1038
343 MetGluLeuLeuAspGlnLeuLysYThrGlyCysValAlleLysGlu 359
1039 ATGGAATTTTGAACAACCTTAATATACATCGGGTGTATTATTAAGAGC 1088
359 rLeuArgLysAsnProProValProGlyGlyPheArgValAlleuLys 376
1089 CTTTCGACTGAATCCCGCATTCAGAGAGGTTTCGGGTGCTGTGAAGA 1138
376 hrPheGluLeuAsnGlyTyrgInleProLysGlyTyrAsnValleTyr 392
1139 CTTTGGATTAATGGATACAGATTCACAGGCTGGAAATGTTATCTAC 1188
393 SerleCysAspThrHisAspValAlaAspValPheProAsnLysGlu 409
1189 AGTATCTGTGACTCATGATGTGGCAGAGATCTTACCAACAAGAGAGA 1238
409 uPheGlnProGluArgPheMetSerLysGlyLeuGlnuAspGlySerArg 426
1239 ATTAAATCCTACCGATTCAGTGTCTCCACCCAGAGATGATCCAGT 1288
426 heAsnTyrlleProPheGlyLysGlySerArgMetCysValGlyLysGlu 442
1289 TCAGCTTCATTCATTCATTGGAGAGGCTTAGAGACTGTGTAGCGCAAGA 1338
443 PheAlaLysValleuLeuLysIlePheLeuValGluLeuThrGlnHisC 459
1339 TTTCCAAAATTTCTTCTCAAAATATTACAGTGGAGCTGCCAGGCAATG 1388
459 sAsnTyrlleuSerAsnGlyProProThrMetLysThrGlyProThr 476
1389 TGACTGGCAGCTTCTAATATGAGACTCTCTACAAATAAACAAGTCCACG 1438
476 leTyProValAspAsnLeuProThrLysPheThrSerTyr 489
1439 TGTATCTGTGGACAATCTCCCTGCAAGATTCCACCATTTTC 1479
seq_name: /cgn2_6/ptodata/2/lna/6B_COMB.seq:us-08-882-164D-31
seq_documentation block:
; Sequence 31, Application US/08882164D
; Patent No. 6306624
; GENERAL INFORMATION:
; APPLICANT: Petkovich, P. Martin, White, Jay A.,
; APPLICANT: Beckett, Barbara R., Jones, Glenville
; TITLE OF INVENTION: Retinoid Metabolizing Protein
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Blake, Cassels & Graydon
; STREET: Box 25, Commerce Court West
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5L 1A9
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 MB storage
; OPERATING SYSTEM: MS-DOS 5.1
; SOFTWARE: WORD PERFECT
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/882,164D
; FILING DATE: June 25, 1997
; PRIOR APPLICATION DATA:
```

```
APPLICATION NUMBER: 08/667,546
FILING DATE: June 21, 1996
APPLICATION NUMBER: 08/724,466
FILING DATE: October 1, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Hunt, John C.
REGISTRATION NUMBER: 36,424
REFERENCE/DOCKET NUMBER: 50767/00010
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 863-4344
TELEFAX: (416) 863-2653
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 1725 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-882-164D-31

alignment_scores:
Quality: 1734.00 Length: 496
Ratio: 4.070 Gaps: 4
Percent Similarity: 85.887 Percent Identity: 67.742

alignment_block:
US-09-668-482-2 x US-08-882-164D-31 ..

Align seg 1/1 to: US-08-882-164D-31 from: 1 to: 1725

1 MetGlyLeuTyThrLeuMetValThrPheLeuCysThrIleValLeuPr 17
|||||
25 ATGGGGCTCCGGCGCTCTGGCCAGTGGCTGTGCACCTTCGTCCTCC 74
|||||
17 OValLeuLeuPheLeuAlaValLysLeuTyProGluMetLeuMetIleA 34
75 GCTGCTCTCTCTCGGGCGCTCAAGCTCTGGGACCTGTACTGTGTGA 124
|||||
34 rGArgValAspProAsnCysArgSerProLeuProProGlyThrMetCly 50
125 GCAGCGCGATCGCAGCTGCCCTCCCTTGGCCCCCGGTACCATGGGC 174
|||||
51 LeuProPheIleGlyGluThrLeuGlnLeuIleLeuGlnArgArgLysPh 67
75 TTCCCATCTTTGGGAAACATGCAATGTGCTTGCAGCGAGGAAGTT 224
|||||
67 eleuArgMetLysArgGlnLysTyrgLysCysIleTyrlLysThrIleuP 84
225 TCTCAGATGAAGCGCAGGAATAGCGCTTCATCTACAAAGCGCATCTGT 274
|||||
84 heGlyAsnProThrValArgValMetGlyAlaAspAsnValArgGlnIle 100
275 TTGGGCGCGCCACGCTCGGGTGTATGGCGCGGATATGTGCGCGCATC 324
|||||
101 LeuLeuGlyLuhLysLysLeuValSerValGlnTrpProAlaSerValAr 117
325 TTGCTGGAGAGCACCGCTGTGTCGTCGACACGCGCGCGTGTGGCG 374
|||||
117 gThrIleLeuGlySerAspThrLeuSerAsnValHisGlyValGlnHisL 134
375 CACCATCTGGCGCGCTGCTGCCCTCCAACTGCAGATTCCTCGACACA 424
|||||
134 ysAsnLysLysAlaIleMetArgAlaPheSerArgAspAlaLeuGlu 150
425 AGCAGCAAAAGAGGTATATGAGCGCTTACGCCCGAGGCACTCCAG 474
|||||
151 HistTyrlleProValIleGlnGlnValLysSerAlaIleGlnGluTr 167
475 TGGTACGTGCTGTCGATCGCTGAGGAGTCAAGCATGTCTGGAGCAGTG 524
|||||
167 pleuGlnLysAspSerCys.....ValleuValTyProGluM 180
|||||
525 GCTA.....AGCTGCGCGAGCGCGGCTCTGTGTACCCCGAGG 565
```

```

180 etylslyslleuMetPheArGlleAlaMetArGlleLeuLeuGlyPheGlu 196
    ::::::::::::::::::::::::::::::::::::::::::::::::::::::
566 TGAACGGCTCATGTTCGGCATCGGCATCGCTCTGGTGGCGAG 615
197 ProGUglnIle.....LysThrAspGluGlnGluLeuValGluAlaPh 211
    |||
616 CCGGGTCCAGCGGGCGGGGAGAGCAGACAGCTCGTGAGGCGCTT 665
211 eGluGluMetIleLysAsnLeuPheSerLeuProIleAspValProPhe 228
    |||
666 CGAGAGATGACCGCAATCTCTCTCTCCATGACGTGACCGCTTTA 715
228 eGlyLeuTyrArgGlyLeuArGlaArAsnPhelIleHisSerLysIle 244
    |||
716 GCGGCTCTACCGGGGGTGAAGGGCGGAACCTATACGCGGCGATC 765
245 GluGluAsnIleArGlyLysIleGlnAsp..AspAspAsnGluAsnG 260
    |||
766 GAGGAGAACATTCGCGCAGATCCGCGGCTTCAGGCTACAGACCGGA 815
260 uGlnLysTyrLysAspAlaLeuGlnLeuIleGluAsnSerArgArgS 277
    |||
816 TGGGGGTGCAGAGAGCGGCTGCGACCTCTGATTGACACTCTGGAGA 865
277 eArSpGluProPheSerLeuGlnAlaMetLysGluAlaAlaThrGluLeu 293
    |||
866 GGGGAGAGCGCTGGATATGACGACATAAAACATCTCAACAGAGCTC 915
294 LeuPheGlyGlyHisGluThrThrAlaSerThrAlaThrSerLeuValMe 310
    |||
916 CTCTTGGGTGATGAACATAACACAGCAGTGTGCGACATCATGATCAC 965
310 rPheLeuGlyLeuAsnThrGluValGlnLysValArGluGluValG 327
    |||
966 TTACTAGACACTACCCACATGCTCTCCAGAAATTCAGAGAAGATTA 1015
327 InGluLysValGluMetGlyMetTyrThrProGlyLysGlyLeuSerMet 343
    |||
1016 AGAGCAAGGGCTTACTTGCAGACCAATCAAGACAAAGATTAGACATG 1065
344 GluLeuLeuAspGlnLeuLysTyrThrGlyLysValIleLysGluThrLe 360
    |||
1066 GAAACTTGGAAAGCTTAATAATCATTTGGGTGTGATTAAGGAGACCT 1115
360 uArgIleAsnProProValProGlyGlyPheArGValAlaLeuLysThrP 377
    |||
1116 GCGATTGATTCCTCGGGTCCAGGAGGGTTGCGGTGCTGGAAGACTT 1165
377 heGluLeuAsnGlyTyrGlnIleProLysGlyTyrAsnValIleTyrSer 393
    |||
1166 TTGAGCTGATGATACCATCCAGATCCCAAGGGCTGGAATGTTTATCACT 1215
394 IleCysAspThrHisAspValAlaAspValPheProAsnLysGluGluPh 410
    |||
1216 ATCTGTGACACCCAGCATGTGCGAGATATCTTCACTAACAGAGGAATT 1265
410 eGlnProGluArgPheMetSerLysGlyLeuGluAspGlySerArgPhea 427
    |||
1266 TAAATCCGACCGCTTATAGTGCCTCATCCAGAGAGATGCTCCGGGTCA 1315
427 snTyrIleProPheGlyGlySerArgMetCysValGlyLysGluPhe 443
    |||
1316 GCTTCATTCATTTGGAGGAGGCTTCGAGGCTGTAGGCAAAAGGTTT 1365
444 AlaLysValLeuLeuLysIlePheLeuValGluLeuThrGlnHisCysAs 460
    |||
1366 GCAGAAATTTCTTAAAGATATTACAGTGGAGCTGCTAGGCACTGTGA 1415
460 nTyrIleLeuSerAsnGlyProProThrMetLysThrGlyProThrIleT 477
    |||
1416 TTGGCAGCTTCTAAATGAGACCTCTCAATGAAGACAGCCCACTGTGT 1465

```

```

477 yrProValAspAsnLeuProThrLysPheThrSerTyr 489
    |||
1466 ACCCTGTGACAAATCTCCCTGCAAGATTCACCTACCTTC 1503
seq_name: /cgn2_6/ptodata/2/lna/6B_COMB.seq:US-08-882-164D-38
seq_documentation_block:
; Sequence 38, Application US/08882164D
; Patent No. 6306624
; GENERAL INFORMATION:
; APPLICANT: Petkovich, P. Martin, White, Jay A.,
; TITLE OF INVENTION: Retinoid Metabolizing Protein
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Blake, Cassels & Graydon
; STREET: Box 25, Commerce Court West
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5L 1A9
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 MB storage
; COMPUTER: COMPAQ, IBM PC compatible
; OPERATING SYSTEM: MS-DOS 5.1
; SOFTWARE: WORD PERFECT
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/882,164D
; FILING DATE: June 25, 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/667,546
; FILING DATE: June 21, 1996
; APPLICATION NUMBER: 08/724,466
; FILING DATE: October 1, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Hunt, John C.
; REGISTRATION NUMBER: 36,424
; REFERENCE/DOCKET NUMBER: 50767/00010
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 863-8344
; TELEFAX: (416) 863-2653
; INFORMATION FOR SEQ ID NO: 38:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4164 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-882-164D-38
alignment_scores:
Quality: 1169.00 Length: 1055
Ratio: 3.005 Gaps: 16
Percent Similarity: 36.872 Percent Identity: 29.858
alignment_block:
US-09-668-482-2 x US-08-882-164D-38
Align seg 1/1 to: US-08-882-164D-38 from: 1 to: 4164
34 ArgArGValAspProAsnCysArgSerProLeuProPro..... 46
    |||
25 AGAGGGGTAGAGAGACAGATGCTCCCGGCGGCCCTCATGGGGGTG 74
47 GlyThrMetGlyLeuProPheIleGlyGluThrLeuGlnLeuIleLeuG 63
    |||
75 CGGCTCTCTCCTCCACCT.....CCCTCTC 100
63 InArGArGlyPheLeuArGMetLysArgGlnLysTyrGlyCysIleTyr 79
    |||
101 ACGGAGAGAAATTCTTGACAGTGAAGCCGAGAAATACGGCTTCATCTAC 150
80 LysThrHisLeuPheGlyAsnProThrValArGValMetGlyAlaAspAs 96

```

```
151 AAGAGCATCTGTTGGCGGCCACGGTGGCGTGAATGGCGCGGATATA 200
96 nValArgInIleLeuLeuGlyGlnHisLysLeuValSerValGlnTPR 113
201 TGTGGCGGCACTTGTGGGAGACCGGTTGGTGTGGTGCAGTGC 250
113 roLaserValArgThrIleLeuGlySerAspThrLeuSerValHis 129
251 CCGGTCGGTGGCGACATCCGCGCGCTGGTGCCTCCCACTGCAC 300
130 GlyValGlnHisLysAsnLysLys..... 138
301 GATTCCTGCACAAAGCAGAAAGAGTGAAGGTGAGCTGCACACTCT 350
138 ..... 138
351 TGGCTGGCAGGAGACCTCATCTATGGCTTGGTCAAGCAAAATAGAT 400
138 ..... 138
401 GCGGGGCGAGGGTAGTCTATGTGGTGGGAGCAGACCTCTATCT 450
138 ..... 138
451 GAGATCACTTACCTTTCTGTAGCAGCTGGTGTAGTCTGGGGGGA 500
138 ..... 138
501 CTGAATTCCTGAAAGGACTCGGAAAGCGAAAGGGGGGCGCTGAG 550
139 ..... 140
551 GAAAGTAGAGATTGTAACACTCTCTGCTGGGGGGTGTCAAGCTAT 600
140 eMeTArgAlaPheSerArgAspAlaLeuGlnHisTyrIleProValIleG 157
601 TATGCAAGCCTTCAGCCGAGGCACTCCAGTCACTGCTGCCGTGATCG 650
157 lngIngluValLysSerAlaIlelngInlurPLeuGlnLysAspSerCys 173
651 CTGAGGAAGTCAGCATGTCTGAGCAGTGGTA.....AGTCGC 691
174 ..... 174
692 GCGCAGCGCGCCCTCTGCTACCCGAGGTGAAGCGCTCATGTCCG 741
186 gIleAlaMetArgIleLeuLeuGlyPheGluProGluGlnIle.....L 201
742 CATGCCATGCGCATCTCTGGCTGGCGAGCCGGTCCAGCGGCGAGC 791
201 yStrAspGluGlnLeuValGlnAlaPheGluGluMetIleLysAsn 217
792 GGGAGGACGACAGCAGCTGCTGGAGGCTTTCGAGAGATGACCCGCAAT 841
218 LeuPheSerLeuProIleAspValProPheSerCylLeuTyrArg.... 232
842 CTCTCTCTCTCCCATTTGACGTGCCCTTAGCGGCTGTACCGGTAAG 891
232 ..... 232
892 GGGCGTTTGGGAGTGGAGTAGGGAACCAAGCTCGGGCATCCGCTCA 941
233 ..... 941
942 CCGCAGCTCTCTCCGCGCTCAGGGCGTGAAGCGCGGAACCTTATACA 991
241 sSerLysIleGluGlnAsnIleArgLysLysIleGlnAsp...AspAspA 257
992 GCGCGCATCGAGAGACATTCGCGCAAGATCCGCGGCTTCAGGCTA 1041
257 snGluAsnGluGlnLysTyrLysAspAlaLeuGlnLeuLeuIleGlnAsn 273
::: ::::
```

```
1042 CAGAGCGGATGGGGTTGCCAAGAGCGCCCTGCAGCTCTGATGAGCAC 1091
274 SerArgArgSerAspGluProPheSerLeuGln..... 284
1092 TCGTGAGAGAGGGAGAGAGGCTGGATATGAGTGAGAACCAATTTCA 1141
284 ..... 284
1142 AAGGTGCCAAGGCGCGGAGTGCCTTGACTTTCCAGACACATTTCTG 1191
284 ..... 284
1192 GGGTCCAAAGCCCTGTCAAGGCCCAAGCTACTTCCAAGTGGGGGGA 1241
284 ..... 284
1242 TGTAGTCTAGAGCTTTTCAACTGTGGGTGTGAGCCCTTCAGGAGC 1291
284 ..... 284
1292 CAAGCAACCTTTCAAGAGGTCGCTTAAGACATCTGCATATCCGATAT 1341
284 ..... 284
1342 TTACATCAAGAAACATACAGTAGCAAAATTAACGTTATGAAGTGAAC 1391
284 ..... 284
1392 AAAGATAAATTTATTCGTTGGGGTCCACACACAGAGAACCGTATTAA 1441
284 ..... 284
1442 AGGGTGCAATGGTCTAGAGAGCTGTGGAAGGGGTGGCTGACATATGG 1491
284 ..... 284
1492 GAAGATCCCAAAAGTTCAAAAGGCAAGGCTCATCTAACAAGTTAAAGCG 1541
284 ..... 284
1542 AAGAGCAGATTAAAGGAGTTTGGCTTTTGTGTTGTGCTTGACTTT 1591
284 ..... 284
1592 CTATGACAAACGAGATTTCACCTTGAAAGTCTCCGTCATATTCACA 1641
284 ..... 284
1642 GGTAGGCTTTGTAAACAGTGTATAAAGCTGACATGTATTAAC 1691
285 ..... 285
1692 TTGCGTTTATCTTAGGCACTAAACAAATCGTCAACAGAGCTCTCTT 1741
295 eGlyGlnHisGluThrAlaSerThrAlaThrSerLeuValMetPheL 312
1742 TGGTGTCATGAAACTACAGCACTGCGAGCTCACTGATCACTTACC 1791
312 eugLysLeuAsnThrGluValGlnLysValArgGluGlnValGlnGlu 328
1792 TAGAGCTTACCCACATGTCTCCAGAAAGTTGAGAAAGATAAAGAGC 1841
329 LysVal..... 330
1842 AAGTAGATGATTCAGAGGTTCCCATTTGCTTAGACATTCCTCTAT 1891
330 ..... 330
1892 TAACCAACACACACACCCCACTGTATATAGTTGCTCGATACACCCAG 1941
330 ..... 330
1942 TACTATACAGTGAAGATCTGAGAGCTAGTGGGAGTGTGGGGAGAGAC 1991
```

```

330 ..... 330
1992 TCACCTGCTGAATTTAAAGGAGTGTGTGACTGGCTCTCTTG 2041
330 ..... 330
2042 GGCAAAATTTGACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCC 2091
330 ..... 330
2092 TCACACACACACACATCACCATTATAGACAGGTTCTCTCTCC 2141
330 ..... 330
2142 CTGACCAAGAACATGATATGTGATAGACCAATAGCTGATCAGGCTC 2191
331 .....Glu..MetGlyMetThrProGlyLys.. 339
2192 GCAGTGTGTGAGGGCTCAGGGTATGACCTTTATATACCTGATAGCA 2241
339 ..... 339
2242 ACATGTCTGATATGGGTTAGGCTGAGAGTGTGAAAGAGAGGCC 2291
339 ..... 339
2292 ATCAGGCGCATCAGCTCTTCCCTTTATCCTCTCCATCAGACGCTTC 2341
339 ..... 339
2342 AGCTTAGTTAACAGGTAGCTCTCTGGGCTGACTTTTGTGGAGTG 2391
339 ..... 339
2392 CCCAGGATCACACTCACTTTTATCTGTTCCATAGGCGTTACTT 2441
340 .....GlyLeuSerMetGluLeuLysPheLeu 349
2442 GCAGAGCAATCAGACAGCAAAAGTATGAGATGAACTTGGCAGAGCTT 2491
350 LysThrGlyCysValIleLysGluThrLeuArgIleAsnProProVa 366
2492 AATATACACTGGGTGTCTATTAAGAGACCCCTGATTAATCTCCGT 2541
366 LProGlyGlyPheArgValAlaLeuLysThrPheGlu..... 378
2542 TCCAGAGAGGTTTGGGTTGCTCTGAAGACTTTTGAAGTGAATGTGAGTG 2591
378 ..... 378
2592 CACCTCTGTCCCCACCCACCCGCTGTCAGTCACTCTGCTATGC 2641
378 ..... 378
2642 TGTGAGCATCAGCTGCCAGAGCACTGGCTCACTGCTTGACAGTGC 2691
378 ..... 378
2692 CTGCTCTCTATGCTAGTGGGAACAAATTTGCTCTCTCTTAAATGCAT 2741
378 ..... 378
2742 CCATGCTAGTAATGACTTTTGTGTTGCAAGCTCAGGGCCGGAGATGTC 2791
379 .....LeuAsn..GlyTyrGlnIleProLys 386
2792 AATCTTAGATTTTGTGTTTAAACAGGATACAGATCCCAAG 2841
387 GLyTPAsnValIleTyrSerIleCysAspThrHisAspValAlaAspVa 403
2842 GCGTGAATGTATTATTACATCTGTGACACCCAGAGTGTGCAATAT 2891

```

```

403 LProProAsnLysGluGluPheGlnProGluArgPheMetSerLysGlyL 420
2892 CTTCATACAGAGAGCAATTTATCCGACCGCTTATAGTCTCTC 2941
420 euGluAspGlySerArgPheAsnTyrIleProPheGlyGlySerArg 436
2942 CAGAGATGCTTCCGGTTCAGCTTCATTCATTTGGAGAGGCTTCGG 2991
437 MetCysValGlyLysGluPheAlaLysValLeuLeuLysIlePheLeuVa 453
2992 AGCTGTAGCAAAAGATTGCAAAAATCTCTTAAGATATTACAGT 3041
453 LgluLeuThrGlnHisCysAsnTrpIleLeuSerAsnGlyProProThrM 470
3042 GGAGCTGGCTAGGCACTGTGATGAGCAGCTTCTTAATGACCTCTCA 3091
470 eLysThrGlyProThrIleTyrProValAspAsnLeuProThrLysPhe 486
3092 TGAAGACAGGCCCATGCTGTACCTGTGACATCTCCCTGCAAGATT 3141
487 ThrSerTyr 489
3142 ACCCACTTC 3150

```

seq_name: /cgn2_6/prodata/2/lna/6B_COMB.seq:us-08-882-164D-36

```

seq_documentation_block:
; Sequence 36, Application US/08882164D
; Patent No. 6306624
; GENERAL INFORMATION:
; APPLICANT: Petkovich, P. Martin, White, Jay A.,
; TITLE OF INVENTION: Retinoid Metabolizing Protein
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Blake, Cassels & Graydon
; STREET: Box 25, Commerce Court West
; City: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5L 1A9
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 MB storage
; COMPUTER: COMPAD, IBM PC compatible
; OPERATING SYSTEM: MS-DOS 5.1
; SOFTWARE: WORD PERFECT
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/882,164D
; FILING DATE: June 25, 1997
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 08/667,546
; FILING DATE: June 21, 1996
; APPLICATION NUMBER: 08/724,466
; FILING DATE: October 1, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Hunt, John C.
; REGISTRATION NUMBER: 36,424
; REFERENCE/DOCKET NUMBER: 50767/00010
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 863-4344
; TELEFAX: (416) 863-2653
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2677 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-882-164D-36

```

```

alignment_scores:
Quality: 792.00      Length: 673
Ratio: 2.891         Gaps: 13

```

Percent Similarity: 40.713 Percent Identity: 31.947

Alignment block:
US-09-668-482-2 x US-08-882-164D-36 ..

Align seg 1/1 to: US-08-882-164D-36 from: 1 to: 2677

```

1 MetGlyLeuTyrThrLeuMetValThrPheLeuCystrHisLeuValLeuP 17
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
676 ATGGGGCTCCGGCGCTGCTGCGCAGTGCGCTCTGCACCTTGCTGCTGCGC 725
17 oValLeuLeuPheLeuAlaIaValLysLeuTrpLysLeuMetIleA 34
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
726 GCTGCTGCTTCTTCTGCTGCGATCAAGCTTGAGGACCTGTACTGCGTGA 775
34 rArgValAspProAsnCystrSerProLeuProGlyThrMetGly 50
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
776 GCGGCGCGGACCGCAGTTGTGTCCTCCATTCGCCCGCGGAGCTATSGS 825
51 LeuProPheIleGlyGluThrLeuGlnLeuIleLeuGlnArgArg.LysP 67
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
826 TTCCCTCTTGTGGGAAACCTTGCAATGNTACTNCAAGTAAAGGAGAG 875
67 heLeuArg..... 69
876 TTGGGCGGGACAGCGCTCTCCCGGAGCCCGCGCGGCTGTGGCTTCT 925
69 ..... 69
926 GCTGAAGTCGGGGTAGGGCCCCCGGGAGGAGCATGTATTGCGGCTAGAG 975
69 ..... 69
976 CAGGCGTCGGCGGAGCGCGGCGCTCCCGGKMYMCTCANGCSCRCWMK 1025
70 ..... MetLysArg 72
1026 TMCCTCCGCTTCTCCACAMAGCGGARSAAKMKCYKMRGATGAAGCGC 1075
73 GlnLysTyrGlyCysIleTyrLysThrHisLeuPheGlyAsnProThiVa 89
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1076 AGGAATATACGGCTTCATCTACAGACGATCTGTTCGGGCGGCCACCGT 1125
89 lArgValMetGlyAlaAspAsnValArgGlnIleLeuLeuGlyLysHisL 106
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1126 ACGGGTATGGCGCGGACATGTGCGCGCATCTGTCTCGAGAGCAGCAC 1175
106 yLeuValSerValGlnTrpProLysSerValArgThrIleLeuGlySer 122
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1176 GGGTGGTGTGGTCTGCTGCGCAGCGTGGTGGCACCATTCTGGAGTCT 1225
123 AspThrLeuSerAsnValHisGlyValGlnHisLysAsnLysLysLys.. 138
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1226 GGGTGGCTCTTAACCTGACAGACTCTCGCAACAGCAGCGCAAGAGGT 1275
138 ..... 138
1276 GGGGCGCAGAGCGAGCGGTGACAGGAGGAGGAGGCCCATTTATGAGCG 1325
138 ..... 138
1326 GAATTCGGCTGATGATGTAGCGCGGGGCTAGCAGCTTGAGGTGGCT 1375
138 ..... 138
1376 AGGACCTCTGCGAGCTCAGGTTAGCTTCCAGCTCGGAGAGAGTGCCAT 1425
138 ..... 138
1426 GTGTCTGGCAGGAGCTGGGGTGTCTGGAAGGAGCGCGGTAGACGAGAG 1475
139 ..... AlaIle 140
  ||| |||

```

```

1476 GGGCGGATGAGGCTTTTAAAGCTGTGCCCTCTCGGAGCTCAGGTGATT 1525
141 MetArgAlaPheSerArgAspAlaLeuGlnHisTyrIleProValIleG1 157
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1526 ATGGGGCTTCAGCGCGGAGGACACTCGAATGCTACGTGCGCGGTATCAC 1575
157 nGlnGluValLysSerAlaIleGlnGluTrpLeuGlnLysAspSerCys. 173
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1576 CGAGGAAGTGGGACAGCAGCTGTGAGCAGTGCGT.....AGCTGCG 1616
174 .....ValLeuValTyrProGluMetLysLysLeuMetPheArg 186
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1617 GCGAGCGGCGCTCTCGTGTCTACCCCGAGGTGAAGCGCTCATGTTCGA 1666
187 lIleAlaMetArgIleLeuLeuGlyPheGluProGluGlnIleLysThrAs 203
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1667 ATCGCATGCGCATCTTACTGGGCTGCGAACC...CAACTGGCGGGGA 1713
203 P.....GluGlnGluLeuValGlnAlaPheGluGluMetIleLysA 217
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1714 CGGGAGCTCCGAGCAGCAGCTTGTGAGGCTTCGAGGAATAGACCGCA 1763
217 snLeuPheSerLeuProIleAspValProPheSerGlyLeuTyrArg... 232
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1764 ATCTCTTCTGCTGCCATTCGACGTGCCCTTCAGCGGCTGTACGGGTA 1813
232 ..... 232
1814 AGGCGGCAAAAGGGCTGCGGACTAGGGGCGGGAGCTGGCGCTGCTCT 1863
233 .....Gly.LeuArgAlaArgAsnPheIle 240
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1864 CACCGCGCGGCTCTGCGGCTCAGGGCATGAAGCGCGGACCTCATTT 1913
241 HisSerLysIleGluGluAsnIleArgLysLysIleGlnAsp...AspAs 256
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1914 CACGGCGCATCGAGCAGACATTCGCGCAAGATCTCGGGCTGCGGCG 1963
256 PAsnGluAsnGlnGlnLysTyrLysAspAlaLeuGlnLeuLeuIleGluA 273
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1964 ATCCGAGGCGGGCCAGGCGCTGCMAAGACGCGCTGAGCTGTGATCAGC 2013
273 snSer..... 274
  ||| |||
2014 ACTGCTGGAGAGGAGGAGAGCGGCTGACATGCAAGTATGACAGCTTC 2063
274 ..... 274
2064 AGACCGAGCACTGGGAGTTGTGCCCTGTGCTTCCAAAGCGCTGTTC 2113
274 ..... 274
2114 TGGGGCCCCCAAGCGCGGCTGGGCGCCAGCTTCTGAGTGGGGGCG 2163
275 .....ArgArgSerAspGluPro... 280
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2164 CGGCTCAGACTACAGTATGGAATCCGAAAGAGAGCTGAGACACCGGT 2213
280 ..... 280
2214 CAGAGAGCTGCGGAAGGGCTCGGAGAACTGGAGCATCCCTTAGC 2263
280 ..... 280
2264 CTTTAMCAGTTTCAAGGAAAGTTGGAATTTGCAAAATGTATATAA 2313
280 ..... 280
2314 GAACCTTGCAATTTTATAAACTAAGACTTTAATCAGAGATTCCGGT 2363
280 ..... 280
2364 AGRGCGGGGTCTACTGCTTACTGCTCCAGCTGAACCTAAAGGAGCTT 2413

```



```

680 TCCTC...TTCTCACCACTTACCGCAAGCCATCCACAGCGGAGAGAG 726
240 LETHSSELVSLLEGLULSNNLEALGYLSYLLLEGLNASPASP 256
727 TGGCGGAGCGGTGACGGTGTGGATGAAAAG...AGGAGAGAGAG 773
257 ASNGLSNGLULSNTYRYSAPALALEUINLEULLEGLUAS 273
774 GAAGAGAGAGCGAGAGAAAAGATATGCTTGGCGGCTTCTT... 818
273 nSerArgArgSerAspGluProPheSerLeuGlnAlaMetIysGluAla 290
819 .....CGCGGAGATGATGATGATTTCCGATGAGAGATTTGCTACTTCT 861
290 LATHRGULLEULPHEGLYGLYHISGLUTRTHALASERTHALATHR 306
862 TGGTGGGCTTACTTGTCCCGCTATGAAACACCTCCACATCATGCT 911
307 SerLeuValMetPheLeuGlyLeuAsnThrGluValAlaGlnIysValAr 323
912 CTCGCCGCAATTTCTCACCGAGACTCTTACGCTTGTCTCAACTCA 961
323 GGLUGLVALGLNGLULSVALGLUMETGLYMETYTRHPRGGLYSG 340
962 GGAAGAG...CATGAAAAGATTAG...GCATGAAAGGATTCGATTA 1005
340 LLeuSerMetGluLeuLeuAspGlnLeuLysTYRTHRGLYCYSAValIle 356
1006 GTCCTGAATGAGATGATTACAGTCATGCCATGCATACACATATGCTGT 1055
357 LysGluThrLeuArgIleAsnProProValProGlyGlyPheArgValAl 373
1056 AATGAGACGTACGAGTGGCTACATCATCGCGGTGTTTTCACACGCGC 1105
373 ALLEULSTHRPHEGLULSNGLYTYRGLNILEPROLYSLYTRPASN 390
1106 AATGACCGAGTTCAGATCAAGGTTTAAATTCCAAAAAGGTGGAAAG 1155
390 ALLELYRSERILECYASPTHRHISAPVALAASPVALPHEPROASN 406
1156 TATTCATCATGTTAGAGCGTTCATTAGACCCCAACCACTTCAAGAT 1205
407 LysGluGluPheGlnProGluArgPheMetSerIysGlyLeuGluAspG 423
1206 GCTGGCACTTCAACCTTGAGATGGCAGACAACTCGGTAAACGACAGG 1255
423 YSERARGPHEASNLYLLEPROPHEGLYGLYLSERARGMETCYVALG 440
1256 CCTTGTAAATGTGTACACCGTTTGGTAGAGGCCAAGGCTATGTCCTG 1305
440 LYLSTGUPHEALALYSVALLEULSYLEPHELEUVALIGLULEUTHR 456
1306 GTTACGAGCTGGCTAGGGTTCACACTCTGTTCTTCCACCGCTAGTG 1355
457 GlnHisCysAsnThrPheLeuSerAsnGlyProProThrMetIysThrG 473
1356 ACAGGCTTACGTGGGT.....CTGCAAGACAGCAAGCAAA 1390
473 YPROTHRILEYRPROVALASPASNLEUPROTHRYSPHETHRSEYTYR 490
1391 GCGGTGTTTCTTCCAACTACAGAAACGCAAGAAACGGTACCCGATCTTG 1440
490 ALARG 491
1441 TGAAG 1445

```

seq_name: /cgn2_6/plodata/2/lna/6A.COMB.seq:US-08-724-466B-11

```

; seq_documentation_block:
; Sequence 11, Application US/08724466B
; Patent No. 6063606
; GENERAL INFORMATION:
; APPLICANT: Petkovich, P. Martin, White, Jay A.,

```

```

? APPLICANT: Beckett, Barbara R., Jones, Glenville
? TITLE OF INVENTION: Retinoid Metabolizing Protein
? NUMBER OF SEQUENCES: 30
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Blake, Cassels & Graydon
? STREET: Box 25, Commerce Court West
? City: Toronto
? ZIP: M5L 1A9
? COUNTRY: Canada
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 Mb storage
? COMPUTER: COMPAO, IBM PC compatible
? OPERATING SYSTEM: MS-DOS 5.1
? SOFTWARE: WORD PERFECT
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/724,466B
? FILING DATE: October 1, 1996
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: 08/667,546
? FILING DATE: June 21, 1996
? ATTORNEY/AGENT INFORMATION:
? NAME: Hunt, John C.
? REGISTRATION NUMBER: 36,424
? REFERENCE/DOCKET NUMBER: 50767/00004
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (416) 863-2653
? TELEFAX: (416) 863-4344
? INFORMATION FOR SEQ ID NO: 11:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 351 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
?
? US-08-724-466B-11

```

alignment_scores:

```

Quality: 458.00 Length: 117
Ratio: 4.404 Gaps: 0
Percent Similarity: 88.889 Percent Identity: 74.359

```

alignment_block:

US-09-668-482-2 x US-08-724-466B-11 ..

Align seg 1/1 to: US-08-724-466B-11 from: 1 to: 351

```

292 GlnLeuLeuPheGlyGlyHisGluThrThAlaserThrAlatrSerle 308
1 GAACTCTCTTTGGAGGAGCAAGAAACACGCGCCAGTCCACGACATCTCT 50
308 uValMetPheLeuGlyLeuAsnThrGluValAlaGlnIysValArgGlu 325
51 GATCACTTACTGGGCTCAACCCACATGTTCTCAGAAAAGTGGAGAAAG 100
325 lUValGlnGluLysValGluMetGlyMetYTRHPRGGLYSGLYSG 341
101 AGCTGAAGAGTAAGGCTTACTTTCGCAAGACATCAAGACAACTTG 150
342 SerMetGluLeuLeuAspGlnLeuLysTYRTHRGLYCYSAValIleYsg 358
151 GACATGAAATTTTGGAAACCTTAATATCATCGGGGTGTATTAAAGA 200
358 uThrLeuArgIleAsnProProValProGlyGlyPheArgValAlaLeu 375
201 GACCTTTCAGACTGATCCCACTGATCCAGGAGGATTTGGGTTGCTGTA 250
375 YSTHRPHEGLULEUASNGLYTYRGLNILEPROLYSGLYTPASVALILE 391
251 AGACTTTGAATTAATGATACAGATTCCCAAGGCTGGAATGTTATC 300
392 TyrSerIleCysAspThrHisAspValAlaAspValAlaPheProAsnIysG 408
301 TACAGTATCTGTGATCATGATGTGGCAGAGATCTTCCACCAACAAGA 350

```

408 u 408
351 A 351

seq_name: /cgn2_6/ptodata/2/1na/6B.COMB.seq:US-08-882-164D-11

```
seq_documentation_block:
; Sequence 11, Application US/08882164D
; Patent No. 6306624
; GENERAL INFORMATION:
; APPLICANT: Petkovich, P. Martin, White, Jay A.,
; APPLICANT: Beckett, Barbara R., Jones, Glenville
; TITLE OF INVENTION: Retinoid Metabolizing Protein
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Blake, Cassels & Graydon
; STREET: Box 25, Commerce Court West
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5L 1A9
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 Mb storage
; COMPUTER: COMPAQ, IBM PC compatible
; OPERATING SYSTEM: MS-DOS 5.1
; SOFTWARE: WORD PERFECT
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/882,164D
; FILING DATE: June 25, 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/667,546
; FILING DATE: June 21, 1996
; APPLICATION NUMBER: 08/724,466
; FILING DATE: October 1, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Hunt, John C.
; REGISTRATION NUMBER: 36,424
; REFERENCE/DOCKET NUMBER: 50767/00010
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 863-4344
; TELEFAX: (416) 863-2653
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 351 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-882-164D-11
```

alignment_scores: Length: 117
Quality: 458.00 Gaps: 0
Ratio: 4.404
Percent Similarity: 88.889 Percent Identity: 74.359

alignment_block:
US-09-668-482-2 x US-08-882-164D-11 ..

Align seg 1/1 to: US-08-882-164D-11 from: 1 to: 351

```
292 gUleuLeuPheGlyGlyHisGluThrAlaSerThrAlaThSerle 308
|||||
1 GAACCTCTCTTGAGAGACGAAACGAGCGCAGTGCAGCCACATCTCT 50
308 uValMetPheLeuGlyLeuAsnThrGluValAlaGlnLysValArgGluG 325
|||||
51 GATCACTTACCTGGGCTCTACCCAGATGTCTCCAGAAAGTGCAGAG 100
325 lUValGlnGlnLysValGluMetGlyMetCysThrProGlyLysGlyLeu 341
|||||
101 AGCTGAAGAGTAAAGGTTACTTATGCAAGAGCAATCAAGACAAGATTG 150
```

```
342 SerMetGluLeuLeuAspGlnLeuLysTyrThrGlyCysValIleLysG 358
|||||
151 GACATGGAATTTTGGACACACTTAATACATCGGTGGTTATTAAAGA 200
358 uThrLeuArgIleAsnProProValProGlyGlyPheArgValAlaLeu 375
|||||
201 GACCCTTCACATGATCCGCCAGTTCAGAGAGGTTGGGTGCTCGA 250
375 yThrPheGluLeuAsnGlyTyrGlnIleProLysGlyTyrPasnValIle 391
|||||
251 AGACTTTGAATTAATGATACAGATTCACAGGCGCTGGAATGTATAC 300
392 TyrSerIleCysAspThrHisAspValAlaAspValPheProAsnLysG 408
|||||
301 TACAGATCTGTGATCTCATGATGTGGCAGAGATCTTACCACAAGGA 350
```

seq_name: /cgn2_6/ptodata/2/1na/6B.COMB.seq:US-08-991-677-3

```
seq_documentation_block:
; Sequence 3, Application US/08991677A
; Patent No. 6252135
; GENERAL INFORMATION:
; APPLICANT: Chiang, Vincent L
; APPLICANT: Carraway, Daniel T
; APPLICANT: Smeltzer, Richard H
; TITLE OF INVENTION: Production of Syringyl Lignin in Gymnosperms
; FILE REFERENCE: 50617
; CURRENT APPLICATION NUMBER: US/08/991,677A
; CURRENT FILING DATE: 1997-12-16
; EARLIER APPLICATION NUMBER: US 60/033,381
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 1883
; TYPE: DNA
; ORGANISM: Liquidambar styraciflua
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (74)..(1606)
US-08-991-677-3
```

alignment_scores: Length: 483
Quality: 343.00 Gaps: 17
Ratio: 1.289
Percent Similarity: 55.072 Percent Identity: 25.466

alignment_block:
US-09-668-482-2 x US-08-991-677-3 ..

Align seg 1/1 to: US-08-991-677-3 from: 1 to: 1883

```
15 ValLeuProValLeuLeuPheLeuAlaValLysLeuTrpGluMetle 31
|||||
128 ATATATACCTTGTCTACTCTTATATGGCCTACTATCT..... 163
31 uMetIleArgValAlaProAsnCysArgSerProLeuProProGlyT 48
|||||
164 .CGGCTTGCCAG.....AGACTACCATACCCACAGGCC 197
48 hMetGlyLeuProPheIleGlyGluThrLeuGlnLeuIleLeuGlnArg 64
|||||
198 CAAGAAGCTTACCGGTATCGAAACATGCTCATATGATCAACTACT 247
65 ArgLysPheLeuArgMetLysArgGlnLysTyrGlyCysIleTyrLys 81
|||||
248 CACCGAGAGCTCGCAAACTCGCAAAACATATAGCGGTCTATTCACT 297
```

```

1156 ATGCGTACTGAAGAAAGTCCTTCGCTCCACCCACCACCATCCACTCCTCC 1205
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
370 heaYgVAlAlaLeuLysThrPheGluLeuAsnGlyTYrGlnIleProLys 386
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1206 TCCACGAGACTCGCCGAGACGCCAGGTCGGCGGCTCTACATCATCTCCGCG 1255
|||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
387 GlyTPraSvAlIleYrSerIleCysAspThrHisAPaValAlaAspVa 403
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::
1256 AAATCGGGGGGATGATCAACACCGCTGGCCATCGCCCGGACAAAGAACT 1305
|||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
403 lPheProAsnLysGluGluPheGlnProGluArgPheMetSerLysLyl 420
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::
1306 GTGGGCGCAGCCACAGATCAGTTTAGCCCTCAGGTTCTCAAAAGACGTG 1355
|||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
420 euGlaSP.....GlySerArgPheAsnTYrIleProPheGlyGly 434
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::
1356 TGCCCGATTTCAAAGGACACACTTCGAGTTCACTTCCTCGGTCAGGT 1405
|||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
435 SerArgMetCysValGlyLysGluPheAlaLysValLeuLysIlePh 451
|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1406 CGTCGCTCTTGCCTTCGATGCAACTGCAGCTACGCGCTAGACGAC 1455
|||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
451 eLeuValGluLeuThrGlnHisCys..AsnTPrlLeuSerAsnGly 466
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::
1456 TGtGGCTCACTCCTT...CACTGTTCACTGTGGAGTTCCGGACGG 1501
|||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
seq_name:/cgn2.6/ptodata/2/lna/5B_COMB.seq:US-08-194-981E-3

seq_documentation_block:
; Sequence 3, Application US/08194981E
; Patient No. 5886157
; GENERAL INFORMATION:
; APPLICANT: GUENEBERICH, F. Peter
; APPLICANT: GÜO, Zuyun
; APPLICANT: SANDHU, Punam
; TITLE OF INVENTION: ELLIZABETH M. J.
; TITLE OF INVENTION: EXPRESSION AND PURIFICATION OF
; TITLE OF INVENTION: HUMAN
; TITLE OF INVENTION: CYTOCHROME P450
; NUMBER OF SEQUENCES: 88
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NEEDLE & ROSENBERG, P.C.
; STREET: Suite 1200, 127 Peachtree Street, NE
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303-1811
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/194,981E
; FILING DATE: February 10, 1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Elizabeth Selby
; REGISTRATION NUMBER: 38,298
; REFERENCE/DOCKET NUMBER: 22000.0022
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 688-9880
; TELEFAX: (404) 688-9880
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1591 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO

```

US-08-194-981E-3

alignment_scores: Length: 517
 Quality: 339.50 Gaps: 22
 Ratio: 1.183 Percent Identity: 26.886
 Percent Similarity: 55.513

alignment_block:
 US-09-668-482-2 x US-08-194-981E-3 ..

Align seg 1/1 to: US-08-194-981E-3 from: 1 to: 1591

```

5 ThrLeuMetValThrPheLeuCysThrIleValLeuProValLeuLeuPhe 21
   ::::::::::::::::::::
7 AGCCYAGTGGTCTGTGCTGTGT.....CTCCATGTTGCTCTCT 47
21 eLeuAlaIaValLysLeuTrpGluMetLeuMetIleArgValAsp 38
   ::::::::::::::::::::
48 CCTTCA.....CTCTGAGA.....CAGA 67
38 roAsnCysArgSerProLeuProProGlyThrMetGlyLeuProPheIle 54
   ::::::::::::::::::::
68 GCTCGGGAGAGAAACTCCCTCTGGCCCACTCTCCCACTGAT 117
55 GlyIuThrLeuGlnLeu...IleLeuGlnArgGlyPheLeuArgme 70
   ::::::::::::::::::::
118 GGAATATCTCCATAGATAGTATTAGAGACATCAGCAATCCTTAAACCA 167
70 LysArgGlnLysTyrGlyCysIleTyrIleThrHisLeuPheGlyAsp 87
   ::::::::::::::::::::
168 TCTTCAGAGTCTATGGCCCTGTCTCTCTCTCTCTCTCTCTCTCTCTCT 217
87 roThrValArgValMetGlyAlaAspAsnValArgGlnIleLeu..... 101
   ::::::::::::::::::::
218 CCATAGTGTGCTCATGATGATGAGACAGTGAAGAGAGCCCTGATTCAT 267
102 LeuGlyGlnHisLysLeuValSerValGlnTrpProAlaSerValArgTh 118
   ::::::::::::::::::::
268 CTGGGAGAGAGATTTTCTGAGAGAGCATTTTCCCTGCTGCTGAGAGAGC 317
118 rIleLeuGlySerAspThrLeuSerAsnValHisGlyValGlnHisLysA 135
   ::::::::::::::::::::
318 TAAAGAGAGA.....TTTGAATTTGTTTCAGCA 346
135 snLysLysLys...AlaIleMetArgAlaPheSerArgAspAlaLeuGlu 150
   ::::::::::::::::::::
347 ATGGAAGAAGAAATGGAAGAGATCCGGCTTCTCTCTCATGACGCTGCGG 396
151 HisTyrIleProValIleGlnGlnIuValLysSerAlaIleGlnGlnIuTr 167
   ::::::::::::::::::::
397 AATTTT.....GGGATGGGGAAGAGACATTCAGACCG 431
167 pleuGlnLysAspSerCysValLeuValTyrProGluMetLysLysLeuM 184
   ::::::::::::::::::::
432 TGTTCAGAGAGAGAGCCGCTGCTGTGGAGAGTGAAGAAAACCAAG 481
184 etPheArgIleAlaMetArgIleLeuLeuGlyPheGluProGluGlnIle 200
   ::::::::::::::::::::
482 CCTACACCTGTGATCCCATCTTCATCTGGGCTGCTCCCTGCAATGTG 531
201 .....LysThrAspGluGln... 205
   ::::::::::::::::::::
532 ATCTGCTCCATTATTTTCCATAACGTTTGTATATAAGATCAGCAATT 581
206 ....GluLeuValGluAlaPheGluMetLys..... 216
   ::::::::::::::::::::
582 TCTTAATTAATGGAAGATTGAATGAACAATCAGATTTTGAGAGCC 631
217 .....AsnLeuPheSerLeuProIleAspValProPhe 227
   ::::::::::::::::::::
632 CTTGATTCAGATCTGCATTAATTTTCTCTATCATGATATAC...TTC 678

```

```

228 SerGlyLeuTyrArgGlyLeuArgAlaArgAsn.....PheIleHisSe 242
   ::::::::::::::::::::
679 CCGGAGACTCACAACAATTA...CTTAAAGCTGCTTTATGAAGA 725
242 rLysIleGluGluAsnIleArgLysLysIleGlnAspAspAsnGlu 259
   ::::::::::::::::::::
726 TTAATTTTGAAGAAAGTAAAGAACCAACCAATCATGACATGACATA 775
259 snGluGlnLysTyrLysAspAlaLeuGlnLeuLeuIleGluAsnSerArg 275
   ::::::::::::::::::::
776 ACCCTCAGACTTATGATTCCTCTCTGATGAATAATGAGAGAGAAAG 825
276 ArgSerAsp...GluProPheSerLeuGlnAlaMetLysGluAlaIaTh 291
   ::::::::::::::::::::
826 CACACCAACCATCAGATTTACTATTGAAGCTTGAAGAAACCTGACGT 875
291 rGluLeuLeuPheGlyGlyHisGluThrThrAlaSerThrAlaThrSerL 308
   ::::::::::::::::::::
876 TGACTTGTGGAGCTGGAGACAGACAGACACAAACCTGAGATATG 925
308 euValMetPheLeuGlyLeuAsnThrGluValValGlnLysValArgGlu 324
   ::::::::::::::::::::
926 CTCTCTCTCTCTCTGACAGCACCAGAGTACACGCTAAAGTCAGAGA 975
325 GluValGlnGlnLysValGluMetGlyMetTyrThrProGlyLysGlyLe 341
   ::::::::::::::::::::
976 GAGATTGAGCTGTGATT.....GGCAGAAACCG 1004
341 uSer.....MetGluLeuLeuAspGlnLeuLysTyrThrGlyCysValI 356
   ::::::::::::::::::::
1005 GAGCCCTGATGCAGACAGACAGACACATGCCCTACACAGATCTGTGG 1054
356 IeLysGluThrLeuArgIleAsnProProValProGlyGlyPheArgVal 372
   ::::::::::::::::::::
1055 TGACAGAGCTCAGAGATGATGATGATCTTCTCCACAGAGCTGCCCAT 1104
373 AlaleuLys...ThrPheGluLeuAsnGlyTyrGlnIleProLysGlyTr 388
   ::::::::::::::::::::
1105 GCAGTGACCTGTGACATTAATTCAGAAATCTCATTTCCCAAGGGCAC 1154
388 PsnValIleTyrSerIleCysAspThr...HisAspValAlaAspValP 404
   ::::::::::::::::::::
1155 AACCATATTAATTTCCCTGACTTCTGTCTCATGACACAAAGAA...T 1201
404 heProAsnLysGluGluPheGlnProGluArgPheMetSerLysGlyLeu 420
   ::::::::::::::::::::
1202 TTCCCAACCAAGAGATGTTTGACCCCTCATCACTTT.....CTG 1239
421 GluAspGlySerArgPheAsn.....TyrIleProPheGlyG 433
   ::::::::::::::::::::
1240 GATGAAGCTGACAAATTTAAGAAAGTAATATCAATGACCTTCTCTCAGC 1289
433 yGlySerArgMetLysValGlyLysGluPheAlaLysValLeuLeuLysI 450
   ::::::::::::::::::::
1290 AGGAAGAGGATTTGTGGAGAGAGCCCTGGCGGATGAGAGCTGTTT 1339
450 IePheLeuValGluLeuThrGlnHisCysAsnTrpIleLeuSerAsnGly 466
   ::::::::::::::::::::
1340 TATCTCTGACCTCATTTTACAGAACTTAACTGAATCTCTGTGTGAC 1389
467 ProProThrMetLysThrGlyProThrIleTyrProValAspAsnLeuPr 483
   ::::::::::::::::::::
1390 CCAAGAACCTTGACACACACTCCAGTTGCAATGATTTGGCTCTGTC 1439
483 o 483
1440 G 1440

```

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-201-118-10
 seq_documentation_block:
 ; Sequence 10 Application US/08201118
 ; Patent No. 5786191

```

GENERAL INFORMATION:
APPLICANT: GOLDSTEIN, JOYCE A.
TITLE OF INVENTION: CLONING AND EXPRESSION OF COMPLEMENTARY
TITLE OF INVENTION: DNAS FOR MULTIPLE MEMBERS OF THE HUMAN CYTOCHROME P450 2C
TITLE OF INVENTION: SUBFAMILY
NUMBER OF SEQUENCES: 44
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend Kourie and Crew
STREET: 379 Lytton Avenue
CITY: Palo Alto
STATE: California
COUNTRY: US
ZIP: 94301
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US-08/201,118
FILING DATE: 22-FEB-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/864,962
FILING DATE: 09-APR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Liebeschuetz, Joe
REGISTRATION NUMBER: 37,505
REFERENCE/DOCKET NUMBER: 15280-192-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 326-2400
TELEFAX: (415) 326-2422
INFORMATION FOR SEQ. ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1852 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-201-118-10

alignment_scores:
Quality: 339.50 Length: 517
Ratio: 1.183 Gaps: 22
Percent Similarity: 55.513 Percent Identity: 26.886

alignment block:
US-09-668-482-2 x US-08-201-118-10 ..

Align seg 1/1 to: US-08-201-118-10 from: 1 to: 1852

5 ThrLeuMetValThrPheLeuGysThrIleValLeuProValLeuLeuPh 21
:::||||:|||||
17 TCCTCTGTGCTCCTGCTGCTGT.....CTCTCAGTTTGGCTTCT 57
:::||||:|||||
21 eLeuAlaIleValLysLeuTrpGluMetLeuMetIleArgValAsp 38
:::||||:|||||
58 CCTTTC.....CTCTGAGA.....CAGA 77
:::||||:|||||
38 rosnGysArgSerProLeuProProGlyThrMetGlyLeuProPhe 54
:::||||:|||||
78 GCCTCGGAGAGGAAAACTCCTCGGCCCACTCCTCCAGTAT 127
:::||||:|||||
55 GlyGluThrLeuGlnLeu...IleLeuGlnArgGlySphLeuArg 70
:::||||:|||||
128 GGAATATCTACAGATAGGTATTAGAGATCAGCAATCTTAACCA 177
:::||||:|||||
70 tLysArgGlnLysTyrGlyCysIleTyrIleHisLeuPheGlyAsp 87
:::||||:|||||
178 TCATCAAGGCTATGAGCCCTGTTCACACTGTATTTTGGCCTGAAC 227
:::||||:|||||
```

```

87 roThrValArgValMetGlyAlaAspAsnValArgGlnIleLeu..... 101
|| ||| :: ||| ::| |||
228 CCATAGTGGTCTGCATGATGAAGCAGTAGAGAACCCCTGATTGAT 277
:::||||:|||||
102 LeuGlyGlnHisLysLeuValSerValGlnTrpProValSerValArg 118
|||||:::|||||
278 CTTCGAGAGAGTTTCTCGAAGAGCATTTTCCCTGCTGCAAGAGC 327
:::||||:|||||
118 rIleLeuGlySerAspThrLeuSerAsnValHisGlyValGlnHisLys 135
:::||||:|||||
328 TAACAGAGA.....TTTGAATTTGTTTCAGCA 356
:::||||:|||||
135 snLysLysLys...AlaIleMetArgAlaPheSerArgAspAlaLeuGln 150
|| ||||| ::||| |||||
357 ATGGAAGAAATGAAAGAGATCCGCGCTTCTCCCTCATGACGCTGCG 406
:::||||:|||||
151 HisTyrIleProValIleGlnGlnValLysSerAlaIleGlnGlnTr 167
:::||||:|||||
407 AATTTT.....GGGATGGGGAAGAGAGCATTTGAGGACCG 441
:::||||:|||||
167 PleuGlnLysAspSerCysValLeuValTyrProGluMetLysLysLeu 184
:::||||:|||||
442 TGTTCAGAGAGAACCCGCTGTTGAGAGAGTTGAGAAAACCAAG 491
:::||||:|||||
184 etPheArgIleAlaMetArgIleLeuLeuGlyPheGluProGlnIle 200
:::||||:|||||
492 CCTCACCTGTGATCCACTTCATCTGGGCTGCTCCCTGCAATGTG 541
:::||||:|||||
201 .....LysThrAspGln..... 205
|||::|
542 ATCTGCTCATTTATTTCCATAAGCTTTTGATTTAAGATCAGCAATT 591
:::||||:|||||
206 ...GluLeuValGlnAlaPheGluGluMetLys..... 216
:::||||:|||||
592 TCTTAACCTTAATGAAAGTTGAATGAAACATCAAGATTTTGAGCACC 641
:::||||:|||||
217 .....AsnLeuPheSerLeuProIleAspValProPhe 227
|| ||||| |||||
642 CCTGATCCAGATCTGCATTAATTTTCTCTATCATGATTAC...TTC 688
:::||||:|||||
228 SerGlyLeuTyrArgGlyLeuArgAlaArgAsn.....PheIleHis 242
|| ::||| ::|||
689 CCGGAACCTCACAAACAATTA...CTTAACAACTGCTTTATGAAGA 735
:::||||:|||||
242 rLysIleGlnGlnAsnIleArgLysLysIleGlnAspAspAsnGln 259
|| ||| |||||:::|||||
736 TTATATTTTGGAAAAAGTAAAGAACACCAAGAAATCATGACATGACA 785
:::||||:|||||
259 snGluGlnLysTyrLysAspAlaLeuGlnLeuIleGlnAsnSerArg 275
|| ||| ::| |||
786 ACCCTCAGACTTTATGATTCCTCTGATGAATAATGAGAGAAAG 835
:::||||:|||||
276 ArgSerAsp...GluProPheSerLeuGlnAlaMetLysGlnAlaIle 291
:::||||:|||||
836 CACAAACCAACCATCTGAATTTACTATTGAAGCTTGAACCACTGACGT 885
:::||||:|||||
291 rGluLeuLeuPheGlyGlnHisGlnTrpThrAlaSerThrAlaIle 308
:::||||:|||||
886 TGACTTTTGGAGCTGGACAGAGACAGACAGACCAACCTGAGATATG 935
:::||||:|||||
308 euValMetPheLeuGlyLeuAsnThrGlnValValGlnLysValArg 324
:::||||:|||||
936 CTCCTCTCTCTCTGTAAGACACCAAGAGTACAGCTAAGTCCAGAA 985
:::||||:|||||
325 GluValGlnGlnLysValGluMetGlyMetTyrThrProGlyLysGly 341
|||:::||||:|||||
986 GAGATTGAACGTGAT.....GGCAGAAACCG 1014
:::||||:|||||
341 user.....MetGluLeuLeuAspGlnLeuLysTyrThrGlyCysVal 356
|| ||| ::|||
1015 GAGCCCTGCATGCAGACAGAGACCACTGCTTACAGATGCTGTGG 1064
:::||||:|||||
356 leLysGlnThrLeuArgIleAsnProProValProGlyGlyPheArgVal 372
:::||||:|||||
```

```

1065 TGACAGAGTCCAGACATGATGACCTCTCCGCCACGACCTGCCCAT 1114
373 AlLeuLys...ThrpheGluLeuAnGlyTYrGlnIleProLysGlyTr 388
1115 GCAGTACCTGTGACATTAATTCAGAACTATCTATCTCCCAAGGGCAG 1164
388 pAsnValIleTySerIleGlySerThr...HisAspValAlaAspValP 404
1165 AACCATATTATTTCCCTGACTCTGTGCTACATGACACAAAGAA...T 1211
404 heProAnLysGluGluProGlnProGluArgPheMetSerLysGlyLeu 420
1212 TTCACCAACCCAGATGTTTGACCTCATCATCTT...CTG 1249
421 GluAspGlySerArgPheAsn...TyrIleProPheGlyG 433
1250 GATGAAAGTGGCAATTTTAAAGAACTAATTAATCTCATGCTTTCTCAGC 1299
433 yGlySerArgMetCysValGlyLysGluPheAlaLysValIleuLysI 450
1300 AGGAAACGATTTGTGTGGAGAGCCCTGCCGCGCATGAGCTGTTT 1349
450 IephLeuValGluLeuThrGlnHisCysAsnTrpIleLeuSerAsnGly 466
1350 TATTCCTGACCTCATTTTACAGAACTTAACTGAATCTCTGTTGAC 1399
467 ProProthMetLysThrGlyProThrIleTyProValAspAsnLeuPr 483
1400 CCAAGAACCTTGACACCACTCCAGTTGTCAATGAGATTGCTCTGTGCC 1449
483 O 483
1450 G 1450
seq.name: /cgn2_6/ptodata/2/lna/5B_COMB.seq:us-08-238-821B-10
seq.documentation_block:
; Sequence 10, Application US/08238821B
; Patent No. 5912120
; GENERAL INFORMATION:
; APPLICANT: GOLDSTEIN, Joyce A.
; APPLICANT: ROMKES-SPARKS, Marjorie
; TITLE OF INVENTION: CLONING, EXPRESSION AND DIAGNOSIS OF HUMAN
; TITLE OF INVENTION: CYTOCHROME P450 2C19: THE PRINCIPAL DETERMINANT OF S-
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: US
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/238,821B
; FILING DATE: 06-MAY-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/201,118
; FILING DATE: 22-FEB-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/864,962
; FILING DATE: 09-APR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Liebeschuetz, Joe

```

```

; REGISTRATION NUMBER: 37,505
; REFERENCE/DOCKET NUMBER: 15280-192110US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 326-2400
; TELEFAX: (650) 326-2422
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1852 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: Region
; LOCATION: 1..10
; OTHER INFORMATION: /note= "Corresponds to positions -10 to-1
; OTHER INFORMATION: for 65 of Figure 2."
US-08-238-821B-10

alignment_scores:
Quality: 339.50 Length: 517
Ratio: 1.183 Gaps: 22
Percent Similarity: 55.513 Percent Identity: 26.886

alignment_block:
US-09-668-482-2 x US-08-238-821B-10 ..

Align seg 1/1 to: US-08-238-821B-10 from: 1 to: 1852
5 ThrLeuMetValThrPheLeuCysThrIleValIleProValLeuLeuPh 21
17 TCTCTGTGTGCTCTGTGCTCTGT...CTCTCATGTTTGCTCTCT 57
21 eLeuAlaIleValLysLeuTrpGluMetLeuMetIleArgValAsp 38
58 CTTTCA.....CTCTGAGA.....CACA 77
38 roAnCysArgSerProLeuProGlyThrMetGlyLeuProPheIle 54
78 GCTCTGGAGAGGAAACTCTCTCTGCGCCACCTCTCCAGATGATT 127
128 GGAATATCTCAGATAGATATTAAAGACATCAGCAAACTTAACCA 177
70 tLysArgGlnLysGlyCysIleTyIleTrpHisLeuPheGlyAsn 87
178 TCTCTCAAGTCTATGGCCCTGTGCTGCTGTATTGGCTGAAAC 227
87 roThrValArgValMetGlyAlaAspAsnValArgGlnIleLeu..... 101
228 CCAATAGTGTGCTCATGAGATGAGAACAGCAAGCCGTGATGAT 277
102 LeuGluHisLysLeuValSerValGlnTrpProAlaSerValArgTh 118
278 CTTCGAGAGGAGTTTCTGGAAGGCAATTTCCCACTGCTGAAGAGC 327
118 rIleLeuGlySerAspThrLeuSerAsnValHisGlyValGlnHisLys 135
328 TTAACAGAGA.....TTTGGAATGTTTTCGCA 356
135 snLysLysLys...AlaIleMetArgAlaPheSerArgAspAlaLeuGlu 150
357 ATGGAAGAAATGAGAGAGATCCGCGGTTTCTCCATGACGCTCCGG 406
151 HisTrpIleProValIleGlnGlnValLysSerAlaIleGlnGlnTr 167
407 AATTT.....GGATGGGGAAGAGAGCATTTGGACCG 441
167 pLeuGlnLysAspSerCysValLeuValTyProGluMetLysLysLeu 184
442 TGTTCAGAGAGAGCCGCTTGTGTGAGAGAGTTGAGAAACCAAG 491

```

184 etpheargilealemetargileuleuenglyphegluproglyuglnle 200
 492 CCTCACCTGTGATCCCACTTTCATCCCTGGCTGCTCCCTGCAATGTG 541
 201LysThrAspGluGln.. 205
 542 ATCTGCTCCATTATTTCCATAAACGTTTGATTATTAAGATCATGACAAAT 591
 206GluLeuValGluAlaPheGluGluMetLeuLys..... 216
 592 TCTTACTTAAATGGAAAAAGTTGATGAAAAACATCAAGATTGTGACGCC 641
 217AsnLeuPheSerLeuProIleAspValProPhe 227
 642 CCTGATTCAGATCTGCAATATATTTCTCTCATCATGATTAC...TTC 688
 228 SerGlyLeuIyrArgGlyLeuArgAlaArgasn.....PheIleHis 242
 689 CCGGGAACCTCAACAAATTA...CTTAAAAAGCTGCTTTATGAAAAAG 735
 242 TlysIleGluGlnAsnIleArgLysIleGlnAspAspAsnGln 259
 736 TTAATTTTGGAAAAAGTAAAGAACACCAAGATCATGATGACATGACAA 785
 259 snGluGlnLysTyrLysAspAlaLeuGlnLeuIleGlnAsnSerArg 275
 786 ACCCTGAGACTTATATGATGCTCTGATGAAAAATGGAAGAAAAAG 835
 276 ArgSerAsp...GluProPheSerLeuGlnAlaMetLysGluAlaIle 291
 836 CACAAACCAACCATCTGATTTACTATGAAAGCTTGGAAAAACATGCAAGT 885
 291 rGluLeuLeuPheGlyGlyHisGluThrThrAlaSerThrAlaThrSerL 308
 886 TGACTTGTGGAGCTGGGACAGACACCAAGCAACACCTGAGATATG 935
 308 euValMetPheLeuGlyLeuAsnThrGlnValAlaGlnLysValArgGln 324
 936 CTCCTCTCTCTCTGTAAGACCAACGAGGTACAGCAATAGTCCAGAA 985
 325 GluValGlnGluLysValGluMetGlyMetLysThrProGlyLysGlyLe 341
 986 GAGATTGACCTGTGATT.....GGCAAGAAACCG 1014
 341 user.....MetGluLeuLeuAspGlnLeuLysTyrThrGlyCysVal 356
 1015 GAGCCCTGATGCAAGACAGACGACCATGCTTACACATATGCTGTGG 1064
 356 LysGluThrLeuArgIleAsnProProValProGlyGlyPheArgVal 372
 1065 TGCACGAGGTGCAGAGATACATTGACCTTCCCCACAGCTGCCCAT 1114
 373 AlaLeuLys...ThrPheGluLeuAsnGlyLysGlnIleProLysGlyTr 388
 1115 GCAGTGACCTGTGACATTAATTCAGAAACATATCTCATTCACAGGCGAC 1164
 388 PAsnValIleTyrSerIleCysAspThr...HisAspValAlaAspValP 404
 1165 AACCATATTAATTTCCCTGACTTCTGTCTACATGACAAACAAAGAA...T 1211
 404 heProAsnLysGluGlnPheGlnProGluArgPheMetSerLysGlyLeu 420
 1212 TTCCCAACCCAGAGATGTTTACCTCATCTACTT.....CTG 1249
 421 GluAspGlySerArgPheAsn.....TyrIleProPheGlyGln 433
 1250 GATGAGAGCTGCAATTTAAAGAAAGTAAATACCTCATGCTTCTCAGC 1299
 433 yGlySerArgMetCysValGlyLysGluPheAlaLysValLeuLeuLysI 450
 1300 AGGAAACGAGATTTGTGGAGAGAGCCCTGGCCGCGCATGAGCTGTTT 1349

450 IepheLeuValGluLeuThrGlnHisCysAsnTrpIleLeuSerAsnGly 466
 1350 TATTCCTGACCTCATTTTACAGAACTTAACTGAAATCTGTGTTGAC 1399
 467 ProProThrMetLysThrGlyProThrIleTyrProValAspAsnLeuPr 483
 1400 CCAAGAACCTTGACACACTCCAGTTGTCATGATTTGCTGTGTGCC 1449
 483 o 483
 1450 G 1450

seq_name: /cgn2/6/ptodata/2/1na/PCTUS_COMB.seq.pct-US95-05744-10

```
seq_documentation_block:
; Sequence 10, Application PC/TUS9505744
; GENERAL INFORMATION:
; APPLICANT: GOLDSTEIN, Joyce A.
; APPLICANT: ROMES-SPARKS, Marjorie
; TITLE OF INVENTION: CLONING, EXPRESSION AND DIAGNOSIS OF HUMAN
; TITLE OF INVENTION: CYTOCHROME P450 2C19: THE PRINCIPAL DETERMINANT
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Townsend and Kourie and Crew
; STREET: 379 Lytton Avenue
; CITY: Palo Alto
; STATE: California
; COUNTRY: US
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/05744
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/238, 821
; FILING DATE: 06-MAY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/201, 118
; FILING DATE: 22-FEB-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/864, 962
; FILING DATE: 09-APR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Dow, Karen B.
; REGISTRATION NUMBER: 29, 684
; REFERENCE/DOCKET NUMBER: 15280-192-1-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 326-2400
; TELEFAX: (415) 326-2422
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1852 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; PCT-US95-05744-10
```

alignment_scores: Quality: 339.50 Length: 517
 Ratio: 1.183 Gaps: 22
 Percent Similarity: 55.513 Percent Identity: 26.886

alignment_block: US-09-668-482-2 x PCT-US95-05744-10 ..

OM of: US-09-668-482-2 to: EST:* out_format : pfs

Date: Nov 6, 2001 2:26 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

```
-MODEL=framed_p2n.model -DEV=xlp
-Q/cgn2.1/USPTO_spool/US0968482/runat_05112001.133353.10873/app_query.fasta.1.1680
-DB=EST -QPM=fastap -SUFFIX=st -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000 -OGAPOP=4.500
-OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FCAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DEL0P=6.000
-DELETE=7.000 -START=1 -MATRIX=biosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
-MAXLEN=2000000000 -USER=US0968482@cgn1.1.5307 -NCPD=6
-ICPD=3 -LONGLOG -NO_XLPHY -WAIT -THREADS=1
```

Search information block:

```
Query: US-09-668-482-2
Query length: 492
Database: EST:*
Database sequences: 10228115
Database length: 431459454
Search time (sec): 2793.150000
```

score list:

Sequence	Strd Orig	ZScore	Escore Len	Documentation
gb_est13:AM174347	+ 1083.00	2243.53	7.6e-116	720
gb_est23:AM167038	+ 997.00	2004.83	7.0e-106	640
gb_est13:AM168223	+ 962.00	1990.82	9.0e-102	662
gb_est13:AM153444	+ 867.00	1786.72	2.2e-90	1028
gb_est13:AM153444	+ 855.00	1761.23	5.2e-89	974
gb_est13:AM153444	+ 804.00	1665.17	1.2e-83	666
gb_est13:AM153444	+ 804.00	1665.17	4.1e-83	893
gb_est13:AM153444	+ 792.00	1637.04	4.6e-82	530
gb_est13:AM153444	+ 765.50	1576.99	1.0e-78	768
gb_est13:AM153444	+ 739.50	1525.07	7.9e-76	615
gb_est13:AM153444	+ 722.00	1487.41	9.9e-74	669
gb_est13:AM153444	+ 698.00	1434.20	9.1e-71	847
gb_est13:AM153444	+ 672.00	1384.49	5.3e-68	571
gb_est13:AM153444	+ 637.00	1313.37	4.9e-64	474
gb_est13:AM153444	+ 633.50	1300.07	1.5e-63	537
gb_est13:AM153444	+ 632.00	1288.35	2.7e-63	597
gb_est13:AM153444	+ 628.00	1284.93	1.2e-62	784
gb_est13:AM153444	+ 624.00	1284.93	1.9e-62	522
gb_est13:AM153444	+ 619.50	1275.10	6.6e-62	539
gb_est13:AM153444	+ 619.50	1274.55	7.1e-62	564
gb_est13:AM153444	+ 611.00	1255.74	7.9e-61	666
gb_est13:AM153444	+ 611.00	1255.74	8.7e-61	666
gb_est13:AM153444	+ 604.00	1243.62	3.8e-60	496
gb_est13:AM153444	+ 595.50	1225.40	3.9e-59	512
gb_est13:AM153444	+ 581.00	1193.84	2.2e-57	563
gb_est13:AM153444	+ 553.00	1137.30	7.1e-54	471
gb_est13:AM153444	+ 550.00	1130.93	3.1e-54	474
gb_est13:AM153444	+ 547.50	1123.98	1.7e-53	545
gb_est13:AM153444	+ 547.00	1124.18	1.7e-53	492
gb_est13:AM153444	+ 510.00	1041.76	6.6e-49	731
gb_est13:AM153444	+ 502.50	1030.30	2.8e-48	509
gb_est13:AM153444	+ 481.00	978.15	2.3e-45	919
gb_est13:AM153444	+ 470.00	958.64	2.8e-44	687
gb_est13:AM153444	+ 462.00	949.58	9.0e-44	365
gb_est13:AM153444	+ 458.00	928.24	1.4e-42	1054
gb_est13:AM153444	+ 457.50	928.58	3.7e-43	415
gb_est13:AM153444	+ 456.00	929.57	1.2e-42	671
gb_est13:AM153444	+ 448.00	915.62	7.0e-42	532
gb_est13:AM153444	+ 436.50	891.06	1.6e-40	552
gb_est13:AM153444	+ 425.00	865.47	4.3e-39	623
gb_est13:AM153444	+ 423.50	857.89	1.1e-38	895

seq_est151:AM766054 + 402.00 816.46 2.3e-36 664 + AM766054 d881f06.y1 Harland
gb_est101:BG487359 - 388.00 788.69 8.2e-35 583 + BG487359 d885d03.x1 Wellcome
gb_est175:BE506442 + 387.00 791.04 6.1e-35 405 + BE506442 db83d11.y1 Wellcome
gb_est100:BG408452 + 385.00 790.03 6.9e-35 312 + BG408452 dd01g02.y1 NICHD XG

seq_name: gb_est143:AM174347

seq documentation block:

```
LOCUS AM174347 720 bp mRNA EST 16-NOV-1999
DEFINITION f142a06.y1 sugano kawakami zebrafish DRB danio rerio cDNA clone
2640274.5 similar to SM:CP26_BRARE P79739 CYTOCHROME P450 26.1;
mRNA sequence.
ACCESSION AM174347 GI:6440295
VERSION AM174347.1
KEYWORDS EST.
SOURCE zebrafish.
ORGANISM Danio rerio
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Rasbortinae; Danio.
1 (bases 1 to 720)
Sugano, S., Kawakami, K., Johnson, S., Li, F., Marra, M., Eddy, S.,
Hillier, L., Clifton, S., Allen, M., Gibbons, M., Jost, S., Kucaba, T.,
Martin, J., Pape, D., Steptoe, M., Underwood, K., Theising, B., Ritter,
E., Bowers, Y., Wylie, T., Waterston, R., and Wilson, R.
Washu zebrafish EST Project 1999
Unpublished (1999)
Other ESTs: f142a06.x1
Contact: S.L. Johnson
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: estek@wustl.edu
Library constructed by Dr. Sumio Sugano and Dr. Koichi Kawakami DNA
Sequencing by: Washington University Genome Sequencing Center
Seq primer: T3 ET from Amersham
High quality sequence stop: 478.
Location/Qualifiers
1..720
/organism="Danio rerio"
/strain="AB"
/db_xref="taxon:7955"
/clone="2640274"
/clone_lib="Sugano Kawakami zebrafish DRB"
/sex="mixed (one male and one female, including
unfertilized eggs)"
/dev_stage="adult"
/lab_host="DH10B (phage resistant)"
/notes="Vector: pME18S-FL3; site:1: DraIII (CACTGTCGTC);
site:2: DraIII (CAACGATGTC); 1st strand cDNA was primed
with an oligo(dT) primer [ATGCGCCCTTTTCTTTTCTTTT];
double-stranded cDNA was ligated to a DraIII adaptor
[CTTGGCGCTGTCG], digested and cloned into distinct DraIII
sites of the pME18S-FL3 vector (5' site CACTGTCG, 3' site
CAACGATGTC). XhoI should be used to isolate the cDNA
insert. Size selection was performed to exclude fragments
<1.5kb. Library constructed by Dr. Sumio Sugano
(University of Tokyo Institute of Medical Science) and
kindly donated by Dr. Koichi Kawakami. Custom primers for
sequencing: 5' end primer CTTCTGCTTAAGAGCTGCG and 3' end
primer CAGCTGCACTGCGACACA."
```

FEATURES

source

1..720
/organism="Danio rerio"
/strain="AB"
/db_xref="taxon:7955"
/clone="2640274"
/clone_lib="Sugano Kawakami zebrafish DRB"
/sex="mixed (one male and one female, including
unfertilized eggs)"
/dev_stage="adult"
/lab_host="DH10B (phage resistant)"
/notes="Vector: pME18S-FL3; site:1: DraIII (CACTGTCGTC);
site:2: DraIII (CAACGATGTC); 1st strand cDNA was primed
with an oligo(dT) primer [ATGCGCCCTTTTCTTTTCTTTT];
double-stranded cDNA was ligated to a DraIII adaptor
[CTTGGCGCTGTCG], digested and cloned into distinct DraIII
sites of the pME18S-FL3 vector (5' site CACTGTCG, 3' site
CAACGATGTC). XhoI should be used to isolate the cDNA
insert. Size selection was performed to exclude fragments
<1.5kb. Library constructed by Dr. Sumio Sugano
(University of Tokyo Institute of Medical Science) and
kindly donated by Dr. Koichi Kawakami. Custom primers for
sequencing: 5' end primer CTTCTGCTTAAGAGCTGCG and 3' end
primer CAGCTGCACTGCGACACA."

BASE COUNT

172 a 176 c 196 g 176 t

alignment_scores:

Quality: 1083.00 Length: 232
Ratio: 4.857 Gaps: 0
Percent Similarity: 96.121 Percent Identity: 90.948

alignment_block:

US-09-668-482-2 x AM174347

112 CGTTTACTCTTCTCCGCCGGTGAAGTGTGGAGATGTTAATGATCC 161
 34 rgarvalaspProasnCyarsArgSerProleupProglyThrmetyl 50
 |||||||
 162 GACGAGTCGATCCGAACAGCAGAACTCCTCTCCGCCAGTACATGGGC 211
 51 leupropheileglyluthleuInleuileuInaArgArglyPh 67
 |||||||
 212 TTGCGGTTCTATTGGAGAAACGCTCCAGCTGATCCCTCCAGAGAAAGAACTT 261
 67 eleuarqmetlysarGlnLysTyrGlyCysIleTyrIsthHisleup 84
 |||||||
 262 TCTGGCGATGAAGACGCAAAATACGGGTGATCTCAAGAGCACTCT 311
 84 heglyAsnProThrValArgValMetGlyAlaAspAsnValArgGlnIle 100
 |||||||
 312 TCGGGAAACCCGACTGTACGGGTGATGGGAGCTGATATGTGAGCAGATT 361
 101 leuileuglyluthleuValSerValGlnTPrProAlaSerValAr 117
 |||||||
 362 CTGCGGGGGAACACAGCGTGTCTTCTCAGTGGCCAGCATCGTAGG 411
 117 gthrIleleuglySerAspThrLeuSerAsnValHisGlyValGlnHisL 134
 |||||||
 412 AACCATCTGGGCTCTGACACCCCTCTCCATGTCCATGAGATTCAACACA 461
 134 ysaAsnLysLysLysAlaIleMetArgAlaPheSerArgAspAlaLeuGlu 150
 |||||||
 462 AAAACAAGAAAAAGCCATTATAGAGCGCTCTCTCGAATGCTCTGGAG 511
 151 HstYrIleProValIleGlnGlnIuValLysSerAlaIleGlnGluTr 167
 |||||||
 512 CACTACATTCCCGTGTATCCAGCAGAGTGAAGAGCGCCATACAGCAATG 561
 167 pleugInLysAspSerCysValleuValTyrProGluMetLysLysLeu 184
 |||||||
 562 GCGCAAAAAGACTCCTGCGTGTGTTATCCAGAAAGAAAGAACTCA 611
 184 ecpheArGillealeMetArgIleleuLeu 193
 |||||||
 612 TGTTCGATAGCTATGAGAAATCTGCTT 640

seq_name: gb_est43:AM128823

seq_documentation_block:

LOCUS AM128823 662 bp mRNA EST 25-OCT-1999
 DEFINITION fe37g11.y1 zebrafish Washu MPIMG EST Danio rerio cDNA 5' similar to
 SW:CP26 BRARE P79739 CYTOCHROME P450 26 ;, mRNA sequence.

ACCESSION AM128823.1 GI:6116737

VERSION AM128823.1
 KEYWORDS EST.
 SOURCE zebrafish.
 ORGANISM Danio rerio

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
 Cypriniformes; Cyprinidae; Rasbora; Danio.

REFERENCE 1 (bases 1 to 662)
 Clark, M., Johnson, S.L., Lehrach, H., Lee, R., Li, F., Maria, M., Eddy,
 S., Hillier, L., Kneub, T., Martin, J., Beck, C., Wylie, T., Underwood,
 K., Stepien, M., Theising, B., Allen, M., Bowers, Y., Person, B.,
 Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schuk, R., Ritter, E.,
 Kohn, S., Shih, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R.,
 and Wilson, R.

Washu zebrafish EST Project 1998
 Unpublished (1998)
 Other ESTs: fe37g11.x1
 Contact: Stephen L. Johnson
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: zbrafish@wustl.edu
 CDNA Library Preparation: Matthew Clark. cDNA Library Arrayed by:

Matthew Clark. DNA Sequencing by: Washington University Genome
 Sequencing Center Clone distribution: Genome Systems, St. Louis,
 Missouri (web address: www.genomesystems.com) (email contact:
 info@genomesystems.com) and Research Genetics, Huntsville, Alabama
 (web address: www.resgen.com) (email contact: info@resgen.com) and
 Ressourcenzentrum Primatendatenbank, Berlin, Germany (web address:
 www.rzpd.de)
 Seq primer: T3 ER from Amersham
 High quality sequence stop: 446.
 Location/Qualifiers

FEATURES

source

1.662

/organism="Danio rerio"

/db_xref="taxon:7950"

/clone_lib="zebrafish Washu MPIMG EST"

/sex="mixed"

/tissue_type="26 somite embryos, adult livers, shield

/stage_embryos"

/lab_host="X11-blue MRF"

/note="Vector: pSPORT1; Site-1: NotI; Site-2: SalI; 1st

strand cDNA was primed with a Not I - oligo(dT)15 primer

[5'pgACTGATCTAGATGCGAGCGGCGCCCTTTTCTTTTCTTTT3'];

double-stranded cDNA was ligated to Sal I adaptors (BRL),

digested with Not I and cloned into the Not I and Sal I

sites of the pSPORT1 vector (BRL). Library was constructed

by Matthew Clark (Lehrach lab; ICRF, London and Max Planck

Institut fuer Molekulare Genetik, Berlin). cDNAs for EST

analysis were selected following oligonucleotide

hybridization fingerprinting of arrayed clones from

zebrafish late somitogenesis (26 ss), adult liver or

embryonic shield stage (5.6 h) libraries. Fingerprint

data were used to computationally cluster cDNAs, and a

single cDNA from each cluster was chosen for sequencing.

In some cases multiple members of the same cluster were

sequenced to assess clustering parameters or single clones

were sequenced additional times to assess quality

control."

BASE COUNT 156 a 174 c 174 g 155 t 3 others
 ORIGIN

alignment_scores: Quality: 962.00 Length: 193
 Ratio: 5.090 Gaps: 0
 Percent Similarity: 97.927 Percent Identity: 97.409

alignment_block:

US-09-668-482-2 x AM128823 ..

Align seg 1/1 to: AM128823 from: 1 to: 662

1 MetGlyLeuTyrThrLeuMetValThrPheLeuCysThrIleValLeuPr 17
 |||||||
 84 ATGGGGCTGTACACCTTATGGTCACTTCTCTGACACCATCGTGTGCC 133
 17 ovalleuLeupheleuAlaIaValLysLeuTrpGluMetLeuMetIleA 34
 |||||||
 134 CGTTTACTCTTCTCCGCCGGGTGAAGTGTGGGAGATGTTAATGATCC 183
 34 rgarvalaspProasnCyarsArgSerProleupProglyThrmetyl 50
 |||||||
 184 GACGAGTCGATCCGAACAGCAGAACTCCTCTACCGCAGTACATGGGC 233
 51 leupropheileglyluthleuInleuileuInaArgArglyPh 67
 |||||||
 234 TTGCGGTTCTATTGGAGAAACGCTCCAGCTGATCCCTCCAGAGAAAGATT 283
 67 eleuarqmetlysarGlnLysTyrGlyCysIleTyrIsthHisleup 84
 |||||||
 284 TCTGGCGATGAAGACGCAAAATACGGGTGATCTCAAGAGCACTCT 333
 84 heglyAsnProThrValArgValMetGlyAlaAspAsnValArgGlnIle 100
 |||||||
 334 TCGGGAAACCCGACTGTACGGGTGATGGGAGCTGATATGTGAGCAGATT 383

```

101 LeuLeuGlyGluHisLysLeuValSerValGlnTrpProAlaSerValAr 117
|||||
384 CTGCTGGGGAAACAGAGCTGGTGTCTGTCTGAGTGGCCAGCATCACTGAG 433
|||||
117 gThrIleLeuGlySerAspThrLeuSerAsnValHisGlyValGlnHisL 134
|||||
434 AACCATTCCTGGGCTCTGACACCCCTCTCCATGTCCATGAGTTCACACAA 483
|||||
134 ySAsnLysLysLysAlaIleMetArGAlaPheSerArGspAlaLeuGlu 150
|||||
484 AAAACAGAAAAAGGCCATTATGAGGGCTCTCTCGAGATGCTCTGGAG 533
|||||
151 HisTyrIleProValIleGlnGlnGluValLysSerAlaIleGlnGlnTr 167
|||||
534 CACTCATTCCTCCGTATCCAGCAGAGGTGAAGAGCGCCATCAGAGAAATG 583
|||||
167 PleuGlnLysAspSerCysValLeuValTyrProGlnMetLysLysLeuM 184
|||||
584 GCTGCAAAAAGACTNCTGGTGTCTGGTTATCCAGANATGAGAAACTCA 633
|||||
184 eTrpArGAlaIleAlaMetArGlyIleLeu 193
|||||
634 TGTNTCGGATACCTATGAGAAATCTGCTT 662
|||||
seq_name: gb_est29:AL532444

seq documentation block:
LOCUS AL532444 1028 bp mRNA EST 13-FEB-2001
DEFINITION AL532444 LTI_NFL001_NBC4 Homo sapiens cDNA clone CSODM014YL13 3
PRIME, mRNA sequence.
ACCESSION AL532444
VERSION AL532444.1 GI:12795937
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 1028)
LI,W.B., Gruber,C., Jessee,J., and Polayes,D.
Full-length cDNA libraries and normalization
Unpublished (2001)
COMMENT
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: segreif@genoscope.cns.fr, Web : www.genoscope.cns.fr.
Location/Qualifiers
1.1028
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CSODM014YL13"
/clone_lib="LTI_NFL001_NBC4"
/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="DH10B"
/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed
by life technologies. Contact: Feng liang life
Technologies, a division of Invitrogen 9800 Medical Center
Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
8371 Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"
BASE COUNT 265 a 244 c 225 g 282 t 12 others
ORIGIN
alignment_scores:
Quality: 867.00 Length: 270
Ratio: 3.905 Gaps: 3
Percent Similarity: 82.222 Percent Identity: 64.444

```

```

alignment_block:
US-09-668-482-2 x AL532444/rev ..
Align seg 1/1 to reverse of: AL532444 from: 1 to: 1028

222 ProIleAspValProPheSerGlyLeuTyrArgGlyLeuAlaArgAs 238
|||||
984 CCCATCGACTGCCCT...TCACGGCTGTACCGGGCATGTAGCGGCAAC 938
|||||
238 nPheIleHisSerLysIleGlnGlnHisnIleArGlyLysIleGlnAsp. 254
|||||
937 CTTAATTACGGCGCGCATGACACAGAAATACGCCGCAAGATCTGGGGC 888
|||||
255 ..AspAspAsnGlnLysnGlnGln.LysTyrLysAspAlaLeuGlnLeu 270
|||||
887 TCGGGGCATCCGAGCGGGCGGCGGCTGCAGAAAGACGCCCTGACGTGT 838
|||||
270 uIleGlnAsnSerArGArgSerAspGlnProPheSerIleGlnAlaMetL 287
|||||
837 GATCGAGCAGCTCGGGAGAGGGAGAGCGGCTGCAGACATGACAGCTAA 788
|||||
287 ySGLuAlaAlaThrGlnLeuLeuPheGlyGlyHisGlyThrThrAlaSer 303
|||||
787 ACCAATCTTCACACCGACTCTCTTTGGAGACAGAAACACAGCGGCAT 738
|||||
304 ThrAlaThrSerLeuValMetPheLeuGlyLeuAsnThrGluValAlaG 320
|||||
737 GCAGCCACATCTCTGTATMACTTACCTGGGGCTTACCCACATGTTCTCCA 688
|||||
320 nLysValArgGlnGluValGlnGlnLysValGlnMetGlyMetTyrThrP 337
|||||
687 GAAAGTGCAGAGAGAGCTGAGAGATGAGGCTTACTTTGCAAGACCAATC 638
|||||
337 roGlyLysGlyLeuSerMetGlnLeuLeuAspGlnLeuLysTyrThrGly 353
|||||
637 AAGACACAAAGTTGGACATGAAATTTTGGACACACTTAATATACATCGGG 588
|||||
354 CysValIleLysGlnThrLeuArGlyLeuAsnProProValProGlyGlyP 370
|||||
587 TGTGTTATTAGGAGAGACCTTCGACTGAATCCCAAGTCCAGGAGGCT 538
|||||
370 eArGValAlaLeuLysThrPheGlnLeuAsnGlyTyrGlnIleProLysG 387
|||||
537 TCGGTTCTCTCGAGACTTTTGAATTAATGATACAGATCCCAAGG 488
|||||
387 lYTPAsnValIleTyrSerIleCysAspThrHisAspValAlaAspVal 403
|||||
487 GCTGGAATGTATCTACAGATATCTGTGATACATGATGAGTGGCAGAGATC 438
|||||
404 PheProAsnLysGlnGluPheGlnProGlnArGpPheMetSerLysGlyLe 420
|||||
437 TTCACCAACAGAGAAATTTATCTCTGACCATTCATGCTGCCACACC 388
|||||
420 uGlnAspGlySerArGpPheAsnTyrIleProPheGlyLysLysSerArG 437
|||||
387 AGAGAGATCATCCAGAGTTCACCCATCATTCATTCGAGAGAGCCTTAGGA 338
|||||
437 eCysValGlyLysGlnPheAlaLysValLeuLeuLysIlePheLeuVal 453
|||||
337 GCTGTGTAGGCAAAAGAAATTTGCAAAATTTCTCAAAATATTATACAGT 288
|||||
454 GluLeuThrGlnHisCysAsnThrIleLeuSerAsnGlyProProThme 470
|||||
287 GAGCTGSCAGAGCATGTSACTGCGMCTTCTAAATGACTCTCCACAT 238
|||||
470 lLysThrGlyProThrIleTyrProValAspAsnLeuProThrLysPhe 487
|||||
237 GAAAWCYAGTCCACACCGTGTATCTGTGACAATCTCCCTCAGAGATCA 188
|||||
487 hTrSerTyr 489
|||||
187 CCATATTC 180

```

```

seq_name: gb_est29:AL539668
seq_documentation_block:
LOCUS AL539668 974 bp mRNA EST 16-FEB-2001
DEFINITION AL539668 LTI_FL013_FBrn1 Homo sapiens cDNA clone CS0DF0361G04 5
prime, mRNA sequence.
ACCESSION AL539668
VERSION AL539668.1 GI:12869097
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 974)
AUTHORS Li, W.-B., Gruber, C., Jessee, J., and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
SOURCE
1..974
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DF0361G04"
/clone_id="LTI_FL013_FBrn1"
/dev_stage="pooled tissue from post conception fetuses (20
week, 24 week and 26 week)"
/lab_host="DH10B"
/note="Organ: Fetal brain; Vector: pCMVSPORT 6; 1st strand
cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-stranded cDNA was digested with Not I
and cloned into the Not I and Eco RV sites of the
pCMVSPORT 6 vector. Library was constructed by Life
Technologies. Contact : Feng Liang Life Technologies, a
division of Invitrogen 9800 Medical Center Drive Rockville
, Maryland 20850, USA Fax : (1) 301 610 8371 Email :
liang@lifestech.com URL :
http://fulllength.invitrogen.com"
BASE COUNT 223 a 269 c 305 g 177 t
ORIGIN
alignment_scores:
Quality: 855.00 Length: 297
Ratio: 3.476 Gaps: 5
Percent Similarity: 82.828 Percent Identity: 63.636
alignment_block:
US-09-668-482-2 x AL539668 ..
Align seg 1/1 to: AL539668 from: 1 to: 974
61 ILEUENGLNARGARGYSPhelAurMetLysArgInLysTyrGlycy 77
::: ::::::::::::::::::::::::::::::::::::::::::::
95 CTGACCCAGGCGAGAGATTCTGCAGATGAACGAGCAATATACGGCTT 144
77 siletylystThrHisLeuPheGlyAsnProThrValArgValMetGlyA 94
||||| ::::::::::::::::::::::::::::::::::::::::::::
145 CATTATACAAACCATCTGTTGGGGGGCCACCGTACGGGTATGGCGC 194
94 laaspaenValArgInLleuLeuGlyGlnHisLysLeuValSerVal 110
||||| ::::::::::::::::::::::::::::::::::::::::::::
195 CGGCAATGTGCGCGCATCTTGTCTCGAGAGAC .CGGCTGTGTGGTGC 243
111 GlnTrpProAlaSerValArgThrIleLeuGlySerAspThrLeuSerAs 127
::: ::::::::::::::::::::::::::::::::::::::::::::
244 CACTGGCCAGCGTGGTGGCCACCATTTCTGGATCTGGCTGCTCTTAA 293
127 nValHisGlyValGlnHisLysAsnLysLysAlaIleMetArgAlaP 144
::: ::::::::::::::::::::::::::::::::::::::::::::

```

```

294 CCTGCAGACTCTCTGCACAGACGCGCAAGAGGTGATTATGCGGCGCT 343
144 heserArgaspAlaLeuGlnHisTyrIleProValIleGlnGlnVal 160
||||| ::::::::::::::::::::::::::::::::::::::::::::
344 TCAGCCCGCGAGGA .CTCGAATGTCTACGTGGCGGAGTACACGAGGAAGTG 392
161 LysSerAlaIleGlnGlnLysTyrLeuGlnLysAspSerCys . 173
||||| ::::::::::::::::::::::::::::::::::::::::::::
393 GCGACAGACTCTGCAGCAGTGGCT . . . . . AGCTGCGCGCAGCGCG 433
174 .ValLeuValTyrProGlnMetLysLysLeuMetPheArgIleAlaMetA 190
||||| ::::::::::::::::::::::::::::::::::::::::::::
434 CCTCTGTGTACCCCGAGGTGAAGCGCTTCATGTCTCGAATGCCATGCG 483
190 rglIleuLeuGlyPheGlnProGlnGlnIleLysThrAsp . . . . . 203
||||| ::::::::::::::::::::::::::::::::::::::::::::
484 GCATCTACTGCTGGCTGCGAACCC . . . . . CAATGCGGCGCGAGCGGACTCC 530
204 GlnGlnLeuValGlnAlaPheGlnGlnMetLysAsnLeuPheS 220
||||| ::::::::::::::::::::::::::::::::::::::::::::
531 GAGCAGCAGCTTGTGAGGCGCTTCGAGGAATGACCCGCAATCTCTCTC 580
220 rLeuProIleAspValProPheSerGlyLeuTyrArgGlyLeuArgAlaA 237
||||| ::::::::::::::::::::::::::::::::::::::::::::
581 GGTGCCATTCGACGCTGCTTACGGGGCTGTACCGGGCATGAGGCGC 630
237 rGAsnPheIleHisSerLysIleGlnGlnAsnIleArgLysLysIleGln 253
||||| ::::::::::::::::::::::::::::::::::::::::::::
631 GGAACTCATTCACGCGCGCATGCGAGCAACATTCGCGCCACAGATCTGC 680
254 Asp . . . . . AspaSpasnGluAsnGlnLysTyrLysAspAlaLeuGlnLe 269
||||| ::::::::::::::::::::::::::::::::::::::::::::
681 GGGCGCGCGCATCCGAGCGCGGCGCGGCGCAAGAGCGGCTGACGCT 730
269 uLeuIleGlnAsnSerArgArgSerAspGlnProPheSerLeuGlnAlaM 286
||||| ::::::::::::::::::::::::::::::::::::::::::::
731 GTTGATTCAGCAGCTCTGTGAGAGGAGCGGCTGACATGAGGCGC 780
286 eLysGlnAlaAlaThrGlnLeuLeuPheGlyGlyHisGlnThrAla 302
||||| ::::::::::::::::::::::::::::::::::::::::::::
781 TAAACCAATCTTCAACCACTCTCTTGGAGAGACAGCAAGACGCGCC 830
303 SerThrAlaThrSerLeuValMetPheLeuGlyLeuAsnThrGlnValA 319
||||| ::::::::::::::::::::::::::::::::::::::::::::
831 ACTGCAGCCACATCTGTATCTACTTACCTGGGCGCTACCCACCATGTCT 880
319 lGlnLysValArgGlnGlnValGlnGlnLysValGlnMetGlyMetYrT 336
||||| ::::::::::::::::::::::::::::::::::::::::::::
881 CCAGAAAGTGCAGAGAGAGCTGAAGAGTAAAGGTTTACTTGTGCAAGACA 930
336 hrProGlyLysGlyLeuSerMetGlnLeuLeuAspGlnLeu 349
::: ::::::::::::::::::::::::::::::::::::::::::::
931 ATCAAGACACAACAAGTTGCAATGGAATTTTGGACACTTA 971
seq_name: gb_est71:BE189825
seq_documentation_block:
LOCUS BE189825 696 bp mRNA EST 07-AUG-2000
DEFINITION db61c05.y1 Wellcome CRC PSK egg Xenopus laevis cDNA clone
IMAGE:3377480 5' similar to gb:gb|AF057566.1|AF057566 Xenopus
laevis retinoid acid converting enzyme (XENOPUS);, mRNA sequence.
ACCESSION BE189825
VERSION BE189825.2 GI:9729548
KEYWORDS EST.
SOURCE African clawed frog.
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
Xenopodinae; Xenopus.
REFERENCE 1 (bases 1 to 696)
AUTHORS Clifton, S., Johnson, S. L., Blumberg, B., Song, J., Hillier, L., Page, D.,
Martin, J., Wylie, T., Underwood, K., Theising, B., Bowers, Y., Person
, B., Gibbons, M., Harvey, N., Ritter, E., Jackson, Y., McCann, R.,

```

TITLE Waterston, R. and Wilson, R.
JOURNAL Washu Xenopus EST project, 1999
COMMENT On Jun 22, 2000 this sequence version replaced g1:8668718.
Other_ESTs: db61c05.x1

Contact: Sandy Clifton, Ph.D.
Washu Xenopus EST project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu

Library constructed by N. Garrett, P. Lemaire, A.M. Zorn, and J.B. Gurdon (Wellcome/CRC Institute). DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: Xenopus clones from this library are available through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov
Seq primer: -40RP from gibco
High quality sequence stop: 483.

FEATURES

source
1..696
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/clone="IMAGE:3377480"
/clone_lib="Wellcome CRC psk egg"
/tissue_type="egg"
/lab_host="DH10B (phage-resistant)"
/note="Vector: pBluescript SK-; Site_1: NotI; Site_2: EcoRI. CDNA were oligo-dT primed and directionally cloned. Library was constructed by N. Garrett, P. Lemaire, A.M. Zorn, and J.B. Gurdon (Wellcome/CRC Institute).
Note: This is a Xenopus Gene Collection (XGC) library."

BASE COUNT 156 a 200 c 202 g 138 t
ORIGIN

alignment_scores:
Quality: 807.00 Length: 235
Ratio: 3.917 Gaps: 3
Percent Similarity: 87.660 Percent Identity: 65.532

alignment_block:

US-09-668-482-2 x BE189825 ..

Align seg 1/1 to: BE189825 from: 1 to: 696

```

41 ArgSerProLeuProGlyThrMetGlyLeuProHeIleGlyGluTh 57
||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2 CGGGACCCACTGCCCCCGGCACTATGGGGCTCTTCGGGGAGAC 51
57 rLeuGlnLeuIleLeuGlnArgArgLysPheLeuArgMetLysArgGln 74
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
52 TCTGCAATGCTGCTGCAGAGCGCAAGTCTCCCAATGATGAAGCGTGA 101
74 yStrArgLysIleTyrLysThrIleLeuPheGlyAsnProThValArg 90
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
102 AGTACGGTGCATCTACAAAGACGATCTGTCGTACGCCACGCGTGC 151
91 ValMetGlyIlaAspAsnValArgGlnIleLeuLeuGlyGlnIleLysLe 107
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:
152 GTACAGGGCGCAGAGAACGTTCCGCAAGATCTATGGGGAGACCAAGCT 201
107 uValSerValGlnTrpProAlaSerValArgThrIleLeuGlySerAsp 124
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
202 GGTCGTGGTGGTGGTGGCGGCTCGGTGCGCAGATCTAGGGCGCGCT 251
124 hrLeuSerAsnValHisGlyValGlnHisLysAsnLysLysAlaIle 140
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
252 GTCTGTCAACCTGCAGACTGTGAGCAAGTACACCAAGAAAGTATT 301
141 MetArgLalAheSerArgAspAlaLeuGlnHisTyrTrpLeuValIleG 157
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
302 GCACAAGCCTTCTCCGAGATGCCCTGCACAATTAAGTGGCGCAGATGA 351

```

```

157 nGlnGluValLysSerAlaIleGlnGluTrpLeuGlnLysAspSerCysV 174
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
352 AGAAGAGTGAGACGCTGTAAACCTGTGGCTGCAGAGCGGCCCTGGC 401
174 alLeuValTyrProGluMetLysLysLeuMetPheArgIleAlaMetArg 190
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
402 TCTGATGTATCCCGCATCAAGCAGCATGTCTCCGCAATTCATGAAG 451
191 IleLeuLeuGlyPheGluProGluGlnIleLysTrpAspGluGlnGlu 206
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
452 CTCCTGCTGGCTGGGATGCCAGCGCATGACAGCACCAACATGAGAGAC 501
207 LeuValGluAlaPheGluGlnMetIleLysAsnLeuPheSerLeuPro 223
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
502 GCTGCTCGAAGCCTTCAGAGAAATGACCCGAATCTCTCTGTGCGCTA 551
223 leAspValProPheSerGlyLeuTyrArgGlyLeuArgAlaArgAsn 239
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
552 TTGATGTGCATGTAGCGGCTCTACCGGGTCTGCGGGCTAGAGACCTT 601
240 IleHisSerLysIleGluGluAsnIleArgLysLysIle.....GlnAs 254
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
602 ATTCAATGGCGCAATGTAGTAAACATCGAAGAGAGAGCTCTAGAGAAC 651
254 pAspAspAsnGluAsnGluGlnLysTyrLysAspAlaLeuGlnLeuLeu 271
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
652 AGACGATTAAT.....TGCCGGGATGCCCTCGCAGCTGCTGA 686
271 leGlu 272
|||:
687 TTGAC 691

```

seq_name: gb_est29:AL532445

seq_documentation_block:

LOCUS AL532445 893 bp mRNA EST 13-FEB-2001
DEFINITION AL532445 LTI_NFL001_NBC4 Homo sapiens cDNA clone CS0DM014T113 5
ACCESSION AL532445
VERSION AL532445.1 GI:12795938
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
1 (bases 1 to 893)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: segre@genoscope.cns.fr, Web : www.genoscope.cns.fr.
Location/Qualifiers
1..893
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DM014T113"
/clone_lib="LTI_NFL001_NBC4"
/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="DH10B"
/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by life technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@litech.com URL : http://fulllength.invitrogen.com"

FEATURES

source

3


```

BASE COUNT      196 a      253 c      275 g      167 t      2 others
ORIGIN

alignment_scores:
  Quality:      804.00      Length:      311
  Ratio:        3.282      Gaps:        9
  Percent Similarity: 78.778      Percent Identity: 61.093

alignment_block:
US-09-668-482-2 x AL532445 ..

Align seg 1/1 to: AL532445 from: 1 to: 893

73  GlnlytYrGlyCysIleTyrLysThrHisLeuPheGlyAsnProThrVa 89
   ::::::::::::::::::::
2  AGGAATATCGGCTTCTACACAGACGATCTGTTGGGCGGCCACCGT 51
   ::::::::::::::::::::
89  lArGValMetGlyAlaAspAsnValArgGlnIleLeuLeuGlyLysL 106
   ::::::::::::::::::::
106  yLeuValSerValGlnThrProAlaSerValArgThrIleLeuGlySer 122
   ::::::::::::::::::::
101  GcCTGcTgTGGTCCACGCGCAGCGTGGCGACCATCTCTGGGATCT 150
   ::::::::::::::::::::
123  AspThrLeuSerAsnValHisGlyValGlnHisLysAsnLys.LysLysA 139
   ::::::::::::::::::::
151  GGCTGCTCTCTAACCTGCACGACCTCTCGCACAGACGCGCAACA.... 196
   ::::::::::::::::::::
139  lAlleMetArgAlaPheSerArgAspAlaLeuGlnHisTyrIleProVal 155
   ::::::::::::::::::::
197  .....AGCCGCGAGGACCTCGAATGCTACCTGCGCGTG 229
   ::::::::::::::::::::
156  lIeGlnGlnValLysSerAlaIleGlnGlnTyrPLeuGlnLysAspSe 172
   ::::::::::::::::::::
230  ATCACCGAGGAGTGGCGACGACCTGGAGCGAGTGCTG.....AG 270
   ::::::::::::::::::::
172  rGys.....ValLeuValTyrProGluMetLysLysLeuMetP 185
   ::::::::::::::::::::
271  CCGCGGCGAGCGCGGCTCTGCTACCCCGAGGTGAAGCCCTCATGT 320
   ::::::::::::::::::::
185  heArgIleAlaMetArgIleLeuLeuGlyPheGluProGlnGlnLys 201
   ::::::::::::::::::::
321  TCCGAATCCCATGCGCATCTACTGCGTGGCGAACCC...CACTGGCG 367
   ::::::::::::::::::::
202  ThrAsp.....GlnGlnGlnLeuValGlnAlaPheGlnGlnMetI 215
   ::::::::::::::::::::
368  GCGGACGGGGACTCCGAGCAGCAGCTGTGGAGGCTTGAGGAATGAC 417
   ::::::::::::::::::::
215  eLysAsnLeuPheSerLeuProIleAspValProPheSerGlyLeuTyrA 232
   ::::::::::::::::::::
418  CCCCAATCTCTCTCTGCTGCTCCATCGACCTGCTTACCGGCGCTGACC 467
   ::::::::::::::::::::
232  rGlyLeuArgAlaArgAsnPheIleHisSerLysIleGlnGlnAsnIle 248
   ::::::::::::::::::::
468  GGGGCAATGAAAGGGCGGAACTCTATTCACGGCGCATTCACAGCAAAAT 517
   ::::::::::::::::::::
249  ArgLysLysIleGlnAsp...AspAspAsnGlnAsnGlnGlnLysTyr 264
   ::::::::::::::::::::
518  CCGCGCAAGATCTCGGGGCTGGCGGCGATCCGAGGCGGCGGCGCA 567
   ::::::::::::::::::::
264  sApAlaLeuGlnLeuLeuIleGlnAsnSerArgLysSerAspGlnProp 281
   ::::::::::::::::::::
568  AGACGGGCTGACGCTGTGATCGAGCATCTGGGAGAGGGAGGAGCGGC 617
   ::::::::::::::::::::
281  heSerLeuGlnAlaMetLysGlnAlaAlaThrGlnLeuLeu.PheGlyG 297
   ::::::::::::::::::::
618  TCGACATGCGAGCGCATTAACCATCTTCAACGCAACTCTCTTTGGAG 667
   ::::::::::::::::::::
297  yHisGlnThrThrAlaSerThrAlaThrSerLeuValMetPheLeuGly 314
   ::::::::::::::::::::
668  ACAGAAACACGCGCATGTCAGCACATCTTGATCATTACTGCGGCG 717

```

```

314  euAsnThrGlnValAlaGlnLysValArgGlnGlnValAlaGlnGlnLysVal 330
   ::::::::::::::::::::
718  TCTAACCCACATGTTCTTCAGAAAGTGCAGAAAGCTGCAAGAGGT 767
   ::::::::::::::::::::
331  GluMetGlyMetThrProGlyLysGlySerMetGluLeuLeuAs 347
   ::::::::::::::::::::
768  TTACTTTGCAAGACATCAAGACAAAGATTGACATGGAATTTGGA 817
   ::::::::::::::::::::
818  ACA.CTTAATATACATCGCGGTGTGTATTAGAGACCTTCGACTGATCC 866
   ::::::::::::::::::::
364  rOpProValProGlyGlyPheArgValAla 373
   ::::::::::::::::::::
867  CCACT...TCAGAGGGGTTCGGGGTGTCT 892
   ::::::::::::::::::::
seq_name: gb_est22:AI584636

```

```

seq_documentation_block:
LOCUS      AI584636      530 bp      mRNA      EST      06-APR-1999
DEFINITION fb81e05.y1 Zebrafish washu MPIMG EST Danio rerio cdna 5' similar to
            TR: E79739 P79739 CYTOCHROME P450 ;, mRNA sequence.
ACCESSION  AI584636
VERSION    AI584636.1 GI:4570533
KEYWORDS   EST.
SOURCE     zebrafish.
ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Rasbora; Danio.
REFERENCE  1 (bases 1 to 530)
            Clark,M., Johnson,S.L., Lehnach,H., Lee,R., Li,F., Marra,M., Eddy
            ,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wyllie,T., Underwood
            ,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
            Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,
            Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
            and Wilson,R.
            Washu zebrafish EST Project 1998
            Unpublished (1998)
JOURNAL   Contact: Stephen L. Johnson
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: zbrafish@wustl.edu
COMMENT   CDNA Library Preparation: Matthew Clark. CDNA Library Arrayed by:
            Matthew Clark. DNA Sequencing by: Washington University Genome
            Sequencing Center Clone Distribution: Genome Systems, St. Louis,
            Missouri (web address: www.genomesystems.com) (email contact:
            info@genomesystems.com) and Research Genetics, Huntsville, Alabama
            (web address: www.resgen.com) (email contact: info@resgen.com) and
            ResourceCenter@umr.fraunhofer.de, Berlin, Germany (web address:
            www.rzpd.de)
            Possible reversed clone: similarity on wrong strand
            Seq primer: r3 ET from Amersham
            High quality sequence stop: 493.
            location/Qualifiers

```

FEATURES

```

source
1..530
/organism="Danio rerio"
/db_xref="taxon:7955"
/clone_lib="Zebrafish Washu MPIMG EST"
/sex="mixed"
/tissue_type="26 somite embryos, adult livers, shield
stage embryos"
/lab_host="X1-blue MRP"
/note="Vector: pSPORT1, site.1: NotI; site.2: SalI; 1st
strand cDNA was primed with a Not I - oligo(dT)15 primer
[5'pGATGATGCTAGATCGCGAGCGCGCCCTTTTCTTTTCTTTT3'];
double-stranded cDNA was ligated to Sal I adaptors (BRU),
digested with Not I and cloned into the Not I and Sal I
sites of the pSPORT1 vector (BRU). Library was constructed
by Matthew Clark (Lehnach lab, ICRF, London and Max Planck

```

Institut fuer Molekulare Genetik, Berlin). cDNAs for EST analysis were selected following oligonucleotide hybridization fingerprinting of arrayed clones from zebrafish late somitogenesis (26 ss), adult liver or embryonic shield stage (5.6 h) libraries. Fingerprint data were used to computationally cluster cDNAs, and a single cDNA from each cluster was chosen for sequencing. In some cases multiple members of the same cluster were sequenced to assess clustering parameters or single clones were sequenced additional times to assess quality control.

BASE COUNT 119 a 145 c 136 g 129 t 1 others

alignment_scores:

Quality: 792.00 Length: 154
Ratio: 5.176 Gaps: 0
Percent Similarity: 99.351 Percent Identity: 98.701

alignment_block:

US-09-668-482-2 x A1584636

Align seg 1/1 to: A1584636 from: 1 to: 530

```

1 MetGlyLeuTyrThrLeuMetValThrPheLeuGlyThrIleValLeuPr 17
|||||
69 ATGGGGCTGTACACCTTATGTGTCACCTTTCGACCATGCTGCTAC 118
17 ovalleuLeuPheLeuAlaValLysLeuTrpGlnMetLeuMetIleA 34
|||||
119 CGTTTACTCTTCTCCGCCGCGGTGAAGTTGTGAGATGTTAATGATCC 168
34 TGAAGValAspProAsnCysArgSerProLeuProGlyThrMetGly 50
|||||
169 GACGAGTCGATCCGACCTCGAAGTCTCTACCGCCAGTACCATGGGC 218
5 LeuProheilegIlyLunThrLeuGlnLeuIleLeuGlnArgGlySph 67
|||||
219 TTGGCTTTCATGTGGAACACCTCCAGCTGATCCCTCCAGAGAGAAAGTT 268
67 eleuArgMetLysArgLysIlyTrpGlyCysIleTyrLysThrIleLeu 84
|||||
269 TCGGCGCATGAAGCGCAAGAAATACGGGTGATTCACAGCAACCTCT 318
84 heglAsnProThrValArgValMetGlyAlaAspAsnValArgGlnIle 100
|||||
319 TCGGGAACCCGACTGTCAGGTTGATGGAGCTGATATGTGAGGACATTT 368
100 LeuLeuGlyLunThrLysLeuValSerValGlnTrpProAlaSerValAr 117
|||||
369 CTCTGGGCGACACACAGCTGTGTCTGTCTGAGGCCACATCAGTAG 418
117 gThrIleLeuGlySerAspThrLeuSerAsnValIleHisGlyValGlnHis 134
|||||
419 AACCATCTGGGCTGTACACCTCTTCATGTCATGAGAGTTCAACACA 468
134 ysaAsnLysLysLysAlaIleMetArgAlaPheSerArgAspAlaLeuGlu 150
|||||
469 AAAACAAGAAAAAGCCATATGAGCGCTTCTCTGAGATGCTCTGGAG 518
151 HisTyrIlePro 154
|||||
519 CACTACATTTCC 530

```

seq_name: gb_est43:AM171465

seq_documentation_block:

LOCUS AM171465 765 bp mRNA EST 12-NOV-1999
DEFINITION f142a06.x1 Sugano Kawakami zebrafish DRA Danio rerio cDNA clone
2640274 3' similar to SW:CP26_BRARE P79739 CYTOCHROME P450 26 ;,
mRNA sequence.
ACCESSION AM171465

VERSION AM171465.1 GI:6402990
KEYWORDS EST.
SOURCE zebrafish.
ORGANISM Danio rerio

REFERENCE
AUTHORS Sugano,S., Kawakami,K., Johnson,S., Li,F., Marra,M., Eddy,S.,
Haller,J., Clifton,S., Allen,M., Gibbons,M., Jost,S., Kucaba,T.,
Martin,J., Pape,D., Stepien,M., Underwood,K., Theising,B., Ritter
,E., Bowers,Y., Wylie,T., Waterston,R. and Wilson,R.
Washu Zebrafish EST Project 1999

TITLE JOURNAL
COMMENT Unpublished (1999)
Contact: S.L. Johnson
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu

library constructed by Dr. Sumio Sugano and Dr. Koichi Kawakami DNA
Sequencing by: Washington University Genome Sequencing Center
Seq primer: T7 ET from Amersham
High quality sequence stop: 514.
location/Qualifiers

FEATURES

source

1..765
/organism="Danio rerio"
/strain="AB"
/db_xref="taxon:7955"
/clone_lib="2640274"
/clone_lib="Sugano Kawakami zebrafish DRA"
/sex="mixed (one male and one female, including
unfertilized eggs)"
/dev_stage="adult"
/lab_host="DH10B (phage resistant)"
/note="Vector: pME185-FL3, site_1: DraIII (CACTGTG);
site_2: DraIII (CACCATGTG); 1st strand cDNA was primed
with an oligo(dT) primer [ATGTGGCCCTTTTCTTTTCTTTT];
double-stranded cDNA was ligated to a DraIII adaptor
[TCTTGGCTCTAGG], digested and cloned into distinct DraIII
sites of the pME185-FL3 vector (5' site CACTGTG, 3' site
CACCATGTG). XhoI should be used to isolate the cDNA
insert. Size selection was performed to exclude fragments
<1.5kb. Library constructed by Dr. Sumio Sugano
(University of Tokyo Institute of Medical Science) and
kindly donated by Dr. Koichi Kawakami. Custom primers for
sequencing: 5' end primer CTCTGCTCTTAAAGCTGGC and 3' end
primer CGACCTGAGCTGAGCAACA."
BASE COUNT 225 a 174 c 154 g 212 t
ORIGIN

alignment_scores:

Quality: 765.50 Length: 171
Ratio: 4.845 Gaps: 1
Percent Similarity: 92.398 Percent Identity: 87.135

alignment_block:

US-09-668-482-2 x AM171465/rev ..

Align seg 1/1 to reverse of: AM171465 from: 1 to: 765

```

322 ValArgGluGluValGlnGluLysValGlnMetGlyMetYrThrProGln 338
|||||
765 CTCACACACAGCGGTTCAGACAGGTTAAAT...GCCATATATCTCCGCT 719
338 yLysGlyLeuSerMetGluLeuLeuAspGlnLeuLysTyrThrGlyCysV 355
|||||
718 CATAGGTTTGAAGCATGAGAGCTGTTCACACAGCTGAAGTACATGAGATG 669
355 alileLysGluThrLeuArgIleAsnProProValProGlyGlyPheArg 371
|||||
668 TGAGTAAGAACTCTCAGATCAACCTCTGCTCTGAGAAATTCAGA 620

```

372 ValAlaLeuLysThrPheGluLeuAsnGlyTyrGlnIleProLysGlyTr 388
 |||||:::||||| |||||:::||||| |||||:::||||| |||||:::|||||
 619 GTCCGACTCAGAACCTCTGAATTGATGTACCAAAATCTCTACAGGATG 570
 388 PASValIleTyrSerIleCysAspThrHisAspValAlaAspValPheP 405
 |||||:::||||| |||||:::||||| |||||:::||||| |||||:::|||||
 569 GAACGCTATTACAGCATCTGTGACACGACGACGATGGCCGACGCTTTTC 520
 405 rAsnLysGluGluPheGlnProGluArgPheMetSerLysGlyLeuGlu 421
 |||||:::||||| |||||:::||||| |||||:::||||| |||||:::|||||
 519 CAACAAGAGAGAGTTCACGCCGAGACATTCAGACGAAAGGTCTGAG 470
 422 AspGlySerArgPheAsnTyrIleProPheGlyGlySerArgMetCys 438
 |||||:::||||| |||||:::||||| |||||:::||||| |||||:::|||||
 469 GACGGGTCCAGGTTTAACTACATCCCTTCGAGAGAGATCGAGATGTG 420
 438 sValGlyLysGluPheAlaLysValLeuLeuLysIlePheLeuValGlu 455
 |||||:::||||| |||||:::||||| |||||:::||||| |||||:::|||||
 419 TGTGGGCAGAGGTTCCGCAAGTGTCTACAGATCTTTTGTAGTGAAT 370
 455 eutThrGlnHisCysAsnTyrIleLeuSerAsnGlyProThrMetLys 471
 |||||:::||||| |||||:::||||| |||||:::||||| |||||:::|||||
 369 TAACGACAGATTCGAATGATTCCTCAACGAGACCCCGCAATGAAA 320
 472 ThrGlyProThrIleTyrProValAspAsnLeuProThrLysPheThrSe 488
 |||||:::||||| |||||:::||||| |||||:::||||| |||||:::|||||
 319 ACAGGCCGCGACTGTTTACCAGTGGACATCTCCCTACCAAAATTCACTAG 270
 488 tTyrValArgAsn 492
 |||||:::||||| |||||:::||||| |||||:::||||| |||||:::|||||
 269 TTATGTCAGAAAT 257
 seq_name: gb_est48:AW513600

seq_documentation_block: 618 bp mRNA EST 03-MAR-2000
 LOCUS AW513600 x047b06.x1 NCI_CGAP_Ut1 Homo sapiens cDNA clone IMAGE:2707091 3'
 DEFINITION similar to SW:CP26_HUMAN O43174 CYTOCHROME P450 26 ;, mRNA
 sequence.
 ACCESSION AW513600
 VERSION AW513600.1 GI:7151678
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 618)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-rt@mail.nih.gov
 Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 CNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/URL at:
 Image.lim.gov/image/html/tresources.shtml

FEATURES
 source
 Possible reversed clone: similarity on wrong strand
 Seq primer: -400p from Gibco
 High quality sequence stop: 440.
 Location/Qualifiers

1..618
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2707091"
 /clone_lib="NCI_CGAP_Ut1"
 /tissue_type="well-differentiated endometrial

adenocarcinoma, 7 pooled tumors"
 /lab_host="DH10B"
 /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: SalI;
 Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 1.75 kb. Life Technologies catalog #:
 1158-014"

BASE COUNT 111 a 191 c 199 g 115 t 2 others
 ORIGIN

alignment_scores:
 Quality: 739.50 Length: 207
 Ratio: 4.131 Gaps: 4
 Percent Similarity: 86.473 Percent Identity: 71.498

alignment_block:
 US-09-668-482-2 x AW513600 ..

Align seg 1/1 to: AW513600 from: 1 to: 618

37 AspProAsnGlySerArgSerProLeuProGlyThrMetGlyLeuProph 53
 |||:::||||| |||||:::||||| |||||:::||||| |||||:::|||||
 10 GACCGCAGTGTGCTCCCTCCATGCCCCCGGAGCTATGGGCTTCCCTT 59
 53 eileGlyGluThrLeuGlnLeuIleLeuGlnArgArgLysPheLeuArgM 70
 |||||:::||||| |||||:::||||| |||||:::||||| |||||:::|||||
 60 CTTGGGGGAAACCTTGCAGATGCTACTGCACGCGAGAAATCTCTGAGA 109
 70 eLysArgGlnLysTyrGlyCysIleTyrLysThrHisLeuPheGlyAsn 86
 |||||:::||||| |||||:::||||| |||||:::||||| |||||:::|||||
 110 TGAAGCGCAGGAATACGGCTTCATCAAGACGACATCTGTCGGGCGG 159
 87 ProThrValArgValMetGlyAlaAspAsnValArgGlnIleLeuLeuG 103
 |||||:::||||| |||||:::||||| |||||:::||||| |||||:::|||||
 160 CCCACGCTACGGGTGATGGGCGGACAAATGTGGCGCATCTGTGCTGG 209
 103 YGluHisLysLeuValSerValGlnTyrProAlaSerValArgThrIleL 120
 |||||:::||||| |||||:::||||| |||||:::||||| |||||:::|||||
 210 AGAGCACCAGGCTGTGTGTCGTCACACTGGCCAGCGCTGGTGCCACATTC 259
 120 eugLysSerAspThrLeuSerAsnValHisGlyValGlnHisLysAsnLys 136
 |||||:::||||| |||||:::||||| |||||:::||||| |||||:::|||||
 260 TGGGATCTGCTGCTCTCTAACTGCACGACACTCTCGCACAAACACAGCG 309
 137 LysLysAlaIleMetArgAlaPheSerArgAspAlaLeuGlnHisTyrIL 153
 |||||:::||||| |||||:::||||| |||||:::||||| |||||:::|||||
 310 AAGAAGGTGATTATGCGGGCTTCAGCCGCGAGGCACTCGAATCTACGT 359
 153 eProValIleGlnGlnGluValLysSerAlaIleGlnGluTyrPheLeuG 170
 |||||:::||||| |||||:::||||| |||||:::||||| |||||:::|||||
 360 GCCGCTGATCACCGAGAGAGTGGCGAGCAGCTGGAGAGTGGCTG... 405
 170 ysaSserCys.....ValLeuValTyrProGluMetLysLys 182
 |||||:::||||| |||||:::||||| |||||:::||||| |||||:::|||||
 406AGTCGCGGCGAGCGCGCTCTCTGCTACCCCGAGGTGAAGGCGC 450
 183 LeuMetPheArgIleAlaMetArgIleLeuLeuGlyPheGluProGluG 199
 |||||:::||||| |||||:::||||| |||||:::||||| |||||:::|||||
 451 CTCATGTTCCGATGCGCATGGCTCTCTACTGTCTGCGGAACCC...CA 497
 199 nileLysThrAsp.....GluGlnGluLeuValGluAlaPheGluG 213
 |||||:::||||| |||||:::||||| |||||:::||||| |||||:::|||||
 498 ACTGGCGGGCGAGCGGAGCTCCGACGACGACACTGTGGAGGCTTCGAGG 547
 213 lmetIleLysAsnLeuPheSerLeuProIleAspValProPheSerGly 229
 |||||:::||||| |||||:::||||| |||||:::||||| |||||:::|||||
 548 AATATACCCGCAATCTTCTGCGCTGCGCCATCGACGTCGCTCAGCGGG 597
 230 leuTyrArgGlyLeuArgAla 236
 |||||:::||||| |||||:::||||| |||||:::||||| |||||:::|||||
 598 CTGTACCGGGGCGATGAAGGCN 618

seq_name: gb_est51:AW765767

```

seq_documentation_block: 669 bp mRNA EST 16-FEB-2001
LOCUS AM765767
DEFINITION da77802.y1 Harland stage 19-23 Xenopus laevis cDNA clone
IMAGE:3200906 5' similar to gb:gb|AF057566.1|AF057566 Xenopus
laevis retinoid acid converting enzyme (XENOPUS), mRNA sequence.
ACCESSION AM765767 GI:7697743
VERSION EST.
KEYWORDS African clawed frog.
SOURCE Xenopus laevis
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
Xenopodinae; Xenopus.
REFERENCE 1 (bases 1 to 669)
AUTHORS Clifton, S., Johnson, S.L., Blumberg, B., Song, J., Hillier, L., Pape, D.,
Martin, J., Wylie, T., Underwood, K., Theising, B., Bowers, Y., Person
B., Gibbons, M., Harvey, N., Ritter, E., Jackson, Y., McCann, R.,
Waterston, R. and Wilson, R.
Washu Xenopus EST Project, 1999
Unpublished (1999)
COMMENT Other ESTs: da77802.x1
Contact: Sandy Clifton, Ph.D.
Washu Xenopus EST project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Library constructed by R. Harland, PhD (University of California,
Berkeley)
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: Xenopus clones from this library are available
through the I.M.A.G.E. Consortium/LLNL at:
image.llnl.gov/image/html/resources.shtml
Seq primer: -40R from Gibco
High quality sequence stop: 495.
FEATURES
source
Location/Qualifiers
1..669
/organism="Xenopus laevis"
/db_xref="taxon:8335"
/clone_image="3200906"
/clone_lib="Harland stage 19-23"
/tissue_type="neurala"
/dev_stage="stage 19-23"
/lab_host="DH10B (phage-resistant)"
/Note="Vector: pCS107 (custom); Site 1: NotI; Site 2: SalI
: cDNA made by oligo-dT priming. Library constructed by
Dr. Francesca Mariani in the laboratory of R. Harland,
Ph.D. (University of California, Berkeley). References:
XBF-2 is a transcriptional repressor that converts
ectoderm into neural tissue. Mariani, F.V. Harland, R.M.,
development. 1998 Dec;125(24):5019-31. PMID: 9811586; U1:
99030283; Use of large-scale expression cloning screens in
the xenopus laevis tadpole to identify gene function.
Gramer TC, Liu KT, Mariani FV, Harland RM, Dev Biol.
2000 Dec 15;128(2):197-210. PMID: 11112324; U1: 20564075;
Note: This is a Xenopus Gene Collection (XGC) library."
BASE COUNT 137 a 206 c 187 g 139 t
ORIGIN
alignment_scores:
Quality: 722.00 Length: 204
Ratio: 4.034 Gaps: 0
Percent Similarity: 87.745 Percent Identity: 64.706
alignment_block:
US-09-668-482-2 x AM765767 ..
Align seg 1/1 to: AM765767 from: 1 to: 669
1 MetGlyLeuYrThrLeuMetValThrPheLeuGlyThrIleValLeuPr 17

```

```

|||||
35 ATGATCTGTATATCTGCTGCTACTAGCGCTGTGTGACCTTGCGCTGCC 84
17 oValleuLeuPheLeuAlaAlaVallysLeuTrpGluMetLeuMetIleA 34
85 ATGCTACTCCCTGCTACCGCGTCCCAACTCTGGGATTTGACTGTCTCA 134
34 rGArValAspProAsnGlySerArgSerProLeuProGluThrMetGly 50
135 GCGCAAGATCCACCTGCGCAACCACTGCCCGGCGACTATGGGG 184
51 leuProPheIleGluThrLeuGluLeuLeuGluGlnArgArgLys 67
165 CTGCCCTTCTTGCGGAGACTGTGCAGAACTGTGCGAGGCGCAAGTT 234
67 eLeuArgMetLysArgGlnLysTrpGlyCysIleTrpLysThrHisLeu 84
235 COTCAATGAAAGCGTAGAAAGTAGTGCGATCTCAAGAGCATCTGT 284
84 heGlyAsnProThrValArgValMetGlyAlaAspAsnValArgGlnIle 100
285 TCGTAGCCCTTAGGTTGCGCGCTCAGCGGCGAGAGAGAGTTGCCAGATC 334
101 leuLeuGlyGluHisLysLeuValSerValGlnTrpProAlaSerValAr 117
335 CTATGGGGGAGACACACACTGTGTGCGTGCAGCTGCCGCTCGGTGCG 384
117 gThrIleLeuGlySerAspThrLeuSerAsnValHisGlyValGlnHisL 134
385 CACGATCTTAGGGCGCGCTGCTGTCCACCTGCACGACACTGTGAGCACA 434
134 ySAsnLysLysLysAlaIleMetArgAlaPheSerArgAspAlaLeuGlu 150
435 AGTACACCAAGAACTGATCGCACAGCCTTCTCCGAGATGCTCGAGC 484
151 HisTrpIleProValIleGlnGlnGluValLysSerAlaIleGlnGluTr 167
485 AATTATGTGCCGAGATGAGAGAGAGTGAGAGCGCTGTAAACCTGTG 534
167 pLeuGlnLysAspSerGlyValLeuValTrpProGluMetLysLysLeu 184
535 GCTGAGAGCGCGCCCTGCTGCTGATGTATCCGCCATCAAGGACATGA 584
184 etPheArgIleAlaMetArgIleLeuLeuGlyPheGluProGluGlnIle 200
585 TGATTGCGATGCCATGAGGCTGCTGCTGCCCTGCGATCCCGACGCGATG 634
201 LysThrAspGlu 204
635 GACAGCCACACA 646
seq_name: gb_est29:AL539667
seq_documentation_block: 847 bp mRNA EST 16-FEB-2001
LOCUS AL539667
DEFINITION AL539667 LTL_P1013_Fbrn1 Homo sapiens cDNA clone CS0DP036VG04 3
prime, mRNA sequence.
ACCESSION AL539667
VERSION AL539667.1 GI:12869095
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 847)
AUTHORS Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
COMMENT Genoscope - Centre National de Sequencage
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr,
Location/Qualifiers

```

```

source
1. .847
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DF036YG04"
/clone_lib="FTL.FL013.FBRn1"
/dev_stage="pooled tissue from post conception fetuses (20
week, 24 week and 26 week)"
/lab_host="DH10B"
/Note="Organ: Fetal brain; Vector: pCMVSPORT 6; 1st strand
cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-stranded cDNA was digested with Not I
and cloned into the Not I and Eco RV sites of the
pCMVSPORT 6 vector. Library was constructed by Life
Technologies. Contact : Feng Liang Life Technologies, a
division of Invitrogen 9800 Medical Center Drive Rockville
, Maryland 20850, USA Fax : (1) 301 610 8371 Email :
fliang@lifetech.com URL :
http://fulllength.invitrogen.com"

BASE COUNT      197 a      190 c      204 g      229 t      27 others
ORIGIN

alignment_scores:
Quality:      698.00      Length:      224
Ratio:      3.921      Gaps:      0
Percent Similarity: 79.464      Percent Identity: 60.714

alignment_block:
US-09-668-482-2 x AL539667/rev ..

Align seg 1/1 to reverse of: AL539667 from: 1 to: 847

264 LysAspAlaLeuGlnLeuLeuIleGlnuSsrArGArGSerAspGluPr 280
|||||
840 AAGAGCGCGCTGACGCTKTGTGATCGAGCAGCTGCGAGAGGAGGAGCG 791
|||
280 oPheserLeuGlnAlaMetLysGlnAlaAlaPhrGlnLeuLeuPheGly 297
|||||
790 GCTGGACATGACGACGACTAAGCAATCTTCAACGCAACTCTTTGGAG 741
|||||
297 LysGlnThrThrAlaSerThrAlaThrSerLeuValMetPheLeuGly 313
|||||
740 GACACGAAACACGCGGCGATGCGACGACATCTGTATCACTTACTGCGG 691
|||||
314 LeuAsnThrGluValValGlnLysValArgGluGluValGlnLysVal 330
|||||
690 CTCACCCACACGCTCCGCCAAGAGCGGAGAAAGCTGAAGATGAGG 641
|||||
330 LglMetGlyMetTyrThrProGlyLysGlyLeuSerMetGluLeuVal 347
|||||
640 TTTACTTTGCAAGMCAATCAAGACAAAGTTGGACATGAGAAATTTGG 591
|||||
347 sptInLeuLysTyrThrGlyCysValIleLysGlnThrLeuArgIleAsn 363
|||||
590 AACAACTCAATMCMCGSGGTGTATTAAAGAGACCCCTGACAGCAGAC 541
|||||
364 ProProValProGlyLysPheArgValAlaLeuLysThrPheGluLeuAs 380
|||||
540 CCCCCGMCNCCAGAGAGGCTTCCGAGNNCTCCGAASMCTTCGAATCAAC 491
|||||
380 nGlyTyrGlnIleProLysGlyTyrPasnValIleTyrSerTleCysAspT 397
|||||
490 TGGATCCCMGANCCTCCAGGCGTGAATGTATCTACATCTCTGTGATA 441
|||||
397 hrhAspValAlaAspValPheProAsnLysGlnLugluPheGlnProGlu 413
|||||
440 CCCATGATGCGCAGAGATCTTACCAACAAGAAATTTAACCTTGAC 391
|||||
414 ArgPheMetSerLysGlyLeuGlnAspGlySerArgPheAsnTyrIlePr 430
|||||
390 CGATTCTATGCTGCTCACCCMGAGAGCAGACCCMGCTCAGCCCATVCC 341
|||||
430 opheGlyGlyLysSerArgMetCysValGlyLysGluPheAlaLysValL 447

```

```

|||||
340 ATTTGGAGGAGGCGCTTAGGAGCTGTAGCAAGAAATTCGCAAAATTC 291
|||||
447 euleuLysIlePheLeuValGlnLeuThrGlnIleCysAsnTyrIleLeu 463
|||||
290 CYCCCAAAATATTATTTAGAGTGAGTGSCAGGACATTTGACTGCGCCSBT 241
|||||
464 SerAsnGlyProProThrMetLysThrGlyProThrIleTyrProValAs 480
|||||
240 CTAAATGAGACTCTCTACABGAAACCCAGCCACCGTATCTGTGGA 191
|||||
480 pasnLeuProThrLysPheThr 487
|||||
190 CAATCTCCCTGCAAGATTCACC 169
|||||

seq_name: gb-est101:BG515142

seq_documentation_block: 571 bp mRNA 28-MAR-2001
LOCUS BG515142
DEFINITION dad60f10.y1 Wellcome CRC PCS107 tropicalis egg silurana tropicalis
CDNA clone IMAGE:4464138 5' similar to TR:093323 093323 RETINOIC
ACID CONVERTING ENZYME: ; mRNA sequence.
ACCESSION BG515142 GI:13485799
VERSION BG515142
KEYWORDS EST.
SOURCE western clawed frog.
ORGANISM silurana tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
Xenopodinae; Silurana.
1 (bases 1 to 571)
Clifton, S., Johnson, S.L., Blumberg, B., Song, J., Hillier, L., Pape, D.,
Martin, J., Wylie, T., Underwood, K., Theising, B., Bowers, Y., Person
, B., Gibbons, M., Harvey, N., Ritzer, E., Jackson, Y., McCann, R.,
Waterson, R. and Wilson, R.
Washu Xenopus EST project, 1999
Unpublished (1999)
other_ESTs: dad60f10.x1
Contact: Sandy Clifton, Ph.D.
Washu Xenopus EST project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
Library constructed by A. Zorn and J. Mason (Wellcome/CRC Institute
Center
). DNA sequencing by: Washington University Genome Sequencing
Center
Clone distribution: Xenopus clones from this library are available
through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov
Seq primer: -40RP from gibco
High quality sequence stop: 503.
Location/Qualifiers
1. 571
/organism="Silurana tropicalis"
/db_xref="taxon:8364"
/clone="IMAGE:4464138"
/clone_lib="Wellcome CRC PCS107 tropicalis egg"
/tissue_type="egg"
/lab_host="DH10B (phage-resistant)"
/Note="Vector: PCS107; Site 1: NotI; Site 2: EcoRI; cDNAs
were oligo-dT primed and directionally cloned. Average
insert size 1.5 kb, range 0.5-4 kb. Library constructed by
A. Zorn and J. Mason (Wellcome/CRC Institute)."
```

```

alignment_scores:
Quality:      672.00      Length:      182
Ratio:      4.123      Gaps:      0
Percent Similarity: 89.560      Percent Identity: 67.033

```

alignment_block:
US-09-668-482-2 x BG515142 ..

Align seg 1/1 to: BG515142 from: 1 to: 571

```

1 MetGlyLeuTyrThrLeuMetValThrPheLeuGlyThrIleValLeuLeuP 17
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
25 ATGGATCTATACACTGTGCTCACGAGCGGCTGTGCACCCCTGGCACTGCG 74
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
17 oValLeuLeuPheLeuAlaIleValAllyLeuTrpGluMetLeuMetIleA 34
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
75 AGTACTACTCTGCTGACAGCTGCGCAACCTATGGAGAGTACTGCTAA 124
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
34 rGATValAlaPProAnGysArGSerProLeuProGlyIleThMetGly 50
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
125 GCCGGAAGATGCCGCGTGCAGAAACCTTCCGCCCTGGCACTATGGCG 174
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
51 LeuProPheIleGlyGluThrLeuGluLeuIleLeuGluIleArgArgLysph 67
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
175 CTGCCCTTCTTGGAGAGACTCTGCAATGAGTGTGCAGAGCGCAAGT 224
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
67 eleuArGmetLysArgGlnLysTyrGlyCysIleTyrIleThrIleLeuP 84
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
225 CTTCAAGATTAAAGCAAGAGAGTACGTCGATCTCAAGAGCGATCTAT 274
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
84 heGlyAsnProThrValArgValMetGlyAlaAspAsnValArgIle 100
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
275 TCGGAGAGCCCAAGCGTGGCGCTTACGGGCGCAGAGAACGTTCCGACATC 324
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
101 LeuLeuGlyGluHisLysLeuValSerValGlnTrpProAlaSerValAr 117
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
325 CTACTGGGGGAGCACAAAGCTGTGCTTCACTGCGCGGCGATCTGTCCG 374
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
117 gThrIleLeuGlySerAspThrLeuSerAsnValHisGlyValGlnHisL 134
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
375 CACATCTCTGGGGGCGGCTGTCTGTCCAAACCTGCACGACATCCGAGCA 424
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
134 ySaSnLysLysAlaIleMetArgAlaPheSerArgAspAlaLeuGlu 150
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
425 AGTACACCAAGAGAGTATGCAAGCCTTCCCGGGAAGCCCTCCGC 474
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
151 HisTyrIleProValIleGlnGluValLysSerAlaIleGlnGluTr 167
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
475 AATTAAGTCCCGCTGATGGAAGAAAGTCAAGCTCCGTCACCTGTG 524
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
167 PLeuGlnLysAspSerCysValLeuValTyrProGluMetLysLys 182
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
525 GCTCAAGAGCGACTCGTGTATTTGGTGTATCCGCGCATCAAGCA 570
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

seq_name: gb_est4:AA239785

seq_documentation_block:

LOCUS AA239785 474 bp mRNA EST 03-MAR-1997

DEFINITION mx80a03.r1 Soares mouse NML Mus musculus cDNA clone IMAGE:692620 5' similar to TR:G1001252 G1001252 HYPOTHETICAL 50.6 KD PROTEIN ;

MRNA sequence.

ACCESSION AA239785

VERSION AA239785.1 GI:1863825

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisler,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Thelking,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.

COMMENT The WashU-HHMI Mouse EST Project Unpublished (1996)

CONTACT: Maria M/Mouse EST Project

WashU-HHMI Mouse EST Project

Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:426180
Seq primer: -28ml3 rev2 ET from Amersham
High quality sequence stop: 461.
Location/Qualifiers
1..474
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_image="692620"
/clone_id="Soares mouse NML"
/tissue_type="Liver"
/lab_host="DH10B"
/note="Vector: pT7AD-Pac (Pharmacia) with a modified polylinker. Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5'. TGTTCACCAATCGAAGTGGAGCGCGCGCAATCTTTTCTTTTCTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library constructed and normalized by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 135 a 99 c 118 g 122 t
ORIGIN

alignment_scores:
Quality: 637.00 Length: 158
Ratio: 4.583 Gaps: 0
Percent Similarity: 87.975 Percent Identity: 74.051

alignment_block:

US-09-668-482-2 x AA239785 ..

Align seg 1/1 to: AA239785 from: 1 to: 474

```

320 GlnLysValArgGluGluValGlnGluLysValGluMetGlyMetTyrTh 336
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1 CAGAAAGTTCGAAAGAGATTAAGCAAGAGCGCTTACTTCCAAAGAGCAA 50
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
336 rProGlyLysGlyLeuSerMetGluLeuLeuAspGluLeuLysTyrTrp 353
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
51 TCAAGACACAAAGTTAGACATGAAACTTTGGAACAGCTTAATACACTG 100
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
353 LysCysValIleLysGluThrLeuArgIleAsnProProValProGlyGly 369
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
101 GGTGTGTCATTAAAGAGAGACCTCGGATGATCCTCGGTTCCAGAGAGG 150
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
370 PheArgValAlaLeuLysThrPheGluLeuAsnGlyTyrGlnIleProL 386
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
151 TTTGGGGTTCCTCGAAGACTTTTGAGCTGAATGATACCAAGATCCCA 200
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
386 sGlyTyrAsnValIleTyrSerIleCysAspThrHisAspValAlaAsp 403
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
201 GGCTGGAATTTATTACAGTATCTGACACCCAGATGACAGATTA 250
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
403 alphaProAsnLysGluGluPheGlnProGluArgPheMetSerLysGly 419
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
251 TCTTCACTAACCAAGAGAGAAATTAAATCCGACCGCTTATATGTCCTCAT 300
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
420 LeuGluAspGlySerArgPheAsnTyrIleProPheGlyGlySerArg 436
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
301 CCAAGAGATGCTTCCCGTTCAGCTTCATTCATTTGAGAGAGGCTTCG 350
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
436 gMetCysValGlyLysGluPheAlaLysValLeuLeuLysIlePheLeu 453
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
351 GAGCTGTGTAGGCAAGAGATTTCGAAAATTTCTTAAAGATTATTACAG 400
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

453 atgltleuthrlnhiscysasntripileuSerAsnGlyProProthr 469
 |||||.....
 401 tgcagcttgctgacgactgtgttgacgacttgcAAATGAGACCTCTCA 450
 470 MetLysThrGlyProThrIleThr 477
 |||||.....
 451 ATGAAGACAGGCCCTGCTGTAC 474

seq_name: gb_est72:BE236243

seq_documentation_block:

LOCUS BE236243 537 bp mRNA EST 10-JUL-2000
 DEFINITION 143959 MARC 1Pig Sus scrofa cDNA 5', mRNA sequence.
 ACCESSION BE236243
 VERSION BE236243.1 GI:9020961

KEYWORDS
 SOURCE

ORGANISM

Sus scrofa
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 1 (bases 1 to 537)

REFERENCE

AUTHORS

Stoner, R.T., Heaton, M.P., Grose, W.M., Bennett, G.A., Laegreid, W.W.
 and Keeler, J.W.

TITLE

Design and use of two pooled tissue normalized cDNA libraries for
 EST discovery in swine

JOURNAL
 COMMENT

Unpublished (2000)
 Contact: Smith TPL
 USDA, ARS, US Meat Animal Research Center
 PO Box 166, Clay Center, NE 68933-0166, USA
 Tel: 402 762 4366
 Fax: 402 762 4390

Email: smith@email.marc.usda.gov
 Single pass sequencing. Bases called and alt. trimmed with phred
 v0.980904.e. Vector identified by cross_match with the -minscore 18
 and -minmatch 12 options.

PCR primers

FORWARD: AGGAAACGCTATGACCAT
 BACKWARD: GTTTCCAGTCAGCAGC
 Plate: 85 row: B column: 21

Seq primer: ATTAGGTGACACTATAG.

Location/Qualifiers

1..537

/organism="Sus scrofa"

/db_xref="taxon:9823"

/clone_id="MARC 1Pig"

/tissue_type="pooled"

/lab_host="DH10B"

/note="Vector: pCMV SPORT6; site.1: XbaI; site.2: XhoI;
 library made from pooled tissue from day 11, 13, 15, 20,
 and 30 embryos."

BASE COUNT 96 a 163 c 174 g 104 t
 ORIGIN

alignment_scores:

Quality: 633.50 Length: 177
 Ratio: 4.141 Gaps: 2
 Percent Similarity: 86.441 Percent Identity: 67.797

alignment_block:

US-09-668-482-2 x BE236243 ..

Align seg 1/1 to: BE236243 from: 1 to: 537

25 ValLysLeuTrpGluMetLeuMetIleArgValAspProAsnCysAr 41
 :::::::::::::::
 1 ATCAAACTCGGAGACCTGACTGCGAGACGCGGACCGCAGCTGCAC 50
 41 gSerProLeuProGlyThrMetGlyLeuProPheIleGlyGluThrL 58
 |||||.....
 51 CATTCCCTTGCCTCGAAGACTATGGGCTCCCTCTTTGGGAGACAT 100

58 euGlnLeuIleLeuGlnArgArgLysPheLeuArgMetLysArgGlnLys 74
 |||||.....
 101 TGCAGATGGTCTACACGGAAGAGTTCTGCAGATGAAGCCAGGAAA 150
 75 TyrGlyCysIleTyrLysThrHisLeuPheGlyAsnProThrValArgVa 91
 |||||.....
 151 TACGGTTTCATCTCAAGACGACATCTGTGGGAGGCCACCGGTCGGGT 200
 91 lMetGlyAlaAspAsnValArgGlnIleLeuLeuGlyGlnHisLysLeu 108
 |||||.....
 201 GATGGTGCAGACAACTGGCGGCATCTGTGCGGGAACACCGGCTCG 250
 108 alSerValGlnTrpProAlaSerValArgThrIleLeuGlySerAspThr 124
 |||||.....
 251 TGTGGTCCACTGGCGCGGCTGGTGCACAGATCTGGGCTCTGGCTGC 300
 125 LeuSerAsnValHisGlyValGlnHisLysAsnLysLysAlaIle 141
 |||||.....
 301 CTCTCCAACTCGACAGACTCTCGCACACAGCGCAAGAAGTGATTAT 350
 141 tArgAlaPheSerArgAspAlaLeuGlnHisTyrIleProValIleGln 158
 |||||.....
 351 GCAGGCTTCACGCCGCGCTCCAGTGTCTACGTCCGCTGATCCAG 400
 158 lngLysValLysSerAlaIleGlnGlnTrpLeuGlnLysAspSerCys... 173
 ::|||.....
 401 AGGAAGTGCAGAGTTGCTGGAGCAGTGCTG.....AGTGCAGGA 441
 174ValLeuValTyrProGluMetLysLysLysLeuMetPheArgI 187
 ::|||.....
 442 GAGCGGCGCTCTGTTTACCCAGGTGAAGCGCTCATGTCCCAT 491
 187 eAlaMetArgIleLeuLeuGlyPheGluPro 197
 |||||.....
 492 CGCATGGCATCTCTGTGCGGTGCGAGGCC 522

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2001, 13:25:24 ; Search time 89.59 Seconds
(without alignments)
336.311 Million cell updates/sec

Title: us-09-668-482-4

Percent score: 2611

Sequence: 1 MGUPALIASALCTFVLPDLL.....PTVYPVDNLPARTFHGEI 497

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_0601.*

1:	/SIDSI/gcgcdata/geneseq/geneseq/AA1980.DAT.*
2:	/SIDSI/gcgcdata/geneseq/geneseq/AA1981.DAT.*
3:	/SIDSI/gcgcdata/geneseq/geneseq/AA1982.DAT.*
4:	/SIDSI/gcgcdata/geneseq/geneseq/AA1983.DAT.*
5:	/SIDSI/gcgcdata/geneseq/geneseq/AA1984.DAT.*
6:	/SIDSI/gcgcdata/geneseq/geneseq/AA1985.DAT.*
7:	/SIDSI/gcgcdata/geneseq/geneseq/AA1986.DAT.*
8:	/SIDSI/gcgcdata/geneseq/geneseq/AA1987.DAT.*
9:	/SIDSI/gcgcdata/geneseq/geneseq/AA1988.DAT.*
10:	/SIDSI/gcgcdata/geneseq/geneseq/AA1989.DAT.*
11:	/SIDSI/gcgcdata/geneseq/geneseq/AA1990.DAT.*
12:	/SIDSI/gcgcdata/geneseq/geneseq/AA1991.DAT.*
13:	/SIDSI/gcgcdata/geneseq/geneseq/AA1992.DAT.*
14:	/SIDSI/gcgcdata/geneseq/geneseq/AA1993.DAT.*
15:	/SIDSI/gcgcdata/geneseq/geneseq/AA1994.DAT.*
16:	/SIDSI/gcgcdata/geneseq/geneseq/AA1995.DAT.*
17:	/SIDSI/gcgcdata/geneseq/geneseq/AA1996.DAT.*
18:	/SIDSI/gcgcdata/geneseq/geneseq/AA1997.DAT.*
19:	/SIDSI/gcgcdata/geneseq/geneseq/AA1998.DAT.*
20:	/SIDSI/gcgcdata/geneseq/geneseq/AA1999.DAT.*
21:	/SIDSI/gcgcdata/geneseq/geneseq/AA2000.DAT.*
22:	/SIDSI/gcgcdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2611	100.0	497	19	AAW37734 Human cytochrome P
2	2611	100.0	497	19	AAW44160 Human retinoid met
3	2437	93.3	497	19	AAW37735 Cytochrome P450RAI
4	2437	93.3	497	19	AAW44161 Mouse retinoid met
5	1738	66.6	492	19	AAW37733 Cytochrome zp450RA
6	1738	66.6	492	19	AAW44159 Zebrafish retinoid
7	1113	42.6	216	21	AAI12489 Human PSEC64 prote
8	478	18.3	468	21	AAW20783 Arabidopsis thalia
9	450.5	17.3	461	21	AAW46490 Arabidopsis thalia
10	450.5	17.3	462	21	AAW46489 Arabidopsis thalia
11	450.5	17.3	465	21	AAW46491 Arabidopsis thalia

12	445.5	17.1	430	21	AAW20784 Arabidopsis thalia
13	439	16.8	513	21	AAW07921 A cytochrome P450
14	434	16.6	481	21	AAW30049 Arabidopsis thalia
15	434	16.6	489	21	AAW30048 Arabidopsis thalia
16	433.5	16.6	472	18	AAW27153 Arabidopsis thalia
17	433.5	16.6	472	21	AAW44571 Arabidopsis thalia
18	433.5	16.6	472	21	AAW45022 Arabidopsis thalia
19	433.5	16.6	491	21	AAW45021 Arabidopsis thalia
20	433.5	16.6	492	21	AAW44570 Arabidopsis thalia
21	428	16.4	444	21	AAW44572 Arabidopsis thalia
22	428	16.4	444	21	AAW45023 Arabidopsis thalia
23	412.5	15.8	461	21	AAW11836 Arabidopsis thalia
24	412.5	15.8	462	21	AAW11835 Arabidopsis thalia
25	412.5	15.8	465	21	AAW11834 Arabidopsis thalia
26	406.5	15.6	433	21	AAW30050 Arabidopsis thalia
27	377	14.4	475	21	AAW07676 Amino acid sequenc
28	352	13.5	388	21	AAW20785 Arabidopsis thalia
29	348.5	13.3	511	20	AAW23341 Sweet gum conferyl
30	348.5	13.3	511	21	AAW19694 Amino acid sequenc
31	335.5	12.8	513	22	AAW31008 B. napus F5H polyP
32	335.5	12.8	513	22	AAW48181 Amino acid sequenc
33	333	12.8	520	22	AAW31007 B. napus F5H polyP
34	333	12.8	520	22	AAW48179 B. napus F5H polyP
35	332	12.7	520	22	AAW48180 Human ORFX ORF321
36	327.5	12.5	163	21	AAW40557 Arabidopsis thalia
37	322	12.3	520	19	AAW26640 Arabidopsis ferula
38	322	12.3	520	21	AAW40099 Arabidopsis ferula
39	322	12.3	520	21	AAW15168 Arabidopsis thalia
40	320.5	12.3	520	21	AAW23917 Arabidopsis thalia
41	313.5	12.0	488	21	AAW3337 Arabidopsis thalia
42	313.5	12.0	488	21	AAW51651 Arabidopsis thalia
43	313.5	12.0	527	21	AAW3336 Arabidopsis thalia
44	310	11.9	471	21	AAW33013 Arabidopsis thalia
45	310	11.9	479	21	AAW23012 Arabidopsis thalia

ALIGNMENTS

RESULT 1

AAW37734 ID AAW37734 standard; Protein: 497 AA.

AAW37734; 07-JUN-1998 (first entry)

Human cytochrome P450RAI protein.

Retinoid regulated gene; cytochrome P450 gene; enzyme; oxidative metabolism; P450RAI; retinoic acid; RA; promoter.

Homo sapiens.

WO9749832-A2.

31-DEC-1997.

23-JUN-1997; 97WO-CA00488.

01-OCT-1996; 96US-0724466.

21-JUN-1996; 96US-0667546.

(TOOH) UNITV QUEENS KINGSTON.

Petkovich PW;

WPI: 1998-077193/07.

N-PSDB; AAW09247.

Identifying DNA encoding inducible or suppressible cytochrome P450 - by screening for drugs which reduce the catabolism of retinoic acid, useful in cancer chemotherapy and the treatment of acne and

PT psoriasis
 XX
 PS
 XX Example 4; Pages 58-59; 113pp; English.
 CC This is the amino acid sequence of the human cytochrome P450RA1. Its
 CC expression is dependent on the presence of retinoic acid (RA). The
 CC retinoid-regulated genes such as the inducible cytochrome P450RA1
 CC gene specifically metabolises a derivative of the RA. The cytochrome
 CC P450 gene in general produces enzymes involved in the oxidative
 CC metabolism of endogenous and exogenous compounds. The cytochrome
 CC P450 nucleotide sequence can be used to induce or suppress the
 CC expression of its protein. P450RA1 is highly induced by RA in cell
 CC lines and tissues. This allows for development of a drug screen
 CC using promoters and nucleotide sequences to identify drugs which are
 CC useful for reducing the catabolism of RA.
 XX
 SQ Sequence 497 AA:

Query Match 100.0%; Score 2611; DB 19; Length 497;
 Best Local Similarity 100.0%; Pred. No. 1.5e-253;
 Matches 497; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MGLPILASALCTFVLPILLFLAATKIMDLVCVSGDRSCALPLPPTMGPFPGFTLQM 60
 DB 1 mglpilasalctfvpillflaaklwdlycvsgrdscalp1ppgtmgpfiftglqm 60
 OY 61 VQRRKFLQMKRRKKGFIYKTHLFGRPYRVAGADVRRILLGDRLVSVHMPASVRTL 120
 DB 61 vqrrkflqmkrrkkygfiykthlfgprvrvmgadvrrillgddrlvsvhmpasvrtll 120
 OY 121 GSGCLSNLHDSHKKORRVIMAFSRREALCYVPVITEEVGSSLEQWLSCGERGLLYYPE 180
 DB 121 gsgclsnlhdseshkgrkvimafsrrealcyvpylteevsgslsegwlsgerglllype 180
 OY 181 VKRLMFRIAMRILLCCEPQLAGDSEQOLVAFEMTNLESLPDPVPSGLYRGMKAR 240
 DB 181 vkrlmfriamrillgceplagdsseqqlvafemtrnlslpdpvpsgilyrgmkar 240
 OY 241 NLIHARIEONIRAKICGLRASEAGOGCKDALQLLTEHSMERGEERLDMQALKOSTFELFG 300
 DB 241 nlharieonirakicglraseagcgckdalqlllehswergerldmqalxgstellifg 300
 OY 301 GHETTASATSLITYLGLYPHYLVQKVRBELSKGLICKSNQDNKIDMELLQOLKTYIGCVI 360
 DB 301 ghettasaatslitylgyphylvlqkvrbelksgllicksngdnkldmellqolkyigcvi 360
 OY 361 KETTLNPPVPGGFRVALKTEFLNGYOIPKGMNVITYCTDTHDAEFTPKKEFPNDRPS 420
 DB 361 ketllnppvpgrtrvalkteflngyqipkgnvlyscdthdaefitpkkeefnpdrfs 420
 OY 421 APHPEDASRFSFIPPGGGLRSCGKEPAKILIKIFVELARHCDWOLLNGPPTMKTSPTV 480
 DB 421 aphpedasrfsfipggglrscgkepakilikifvelarhcdwollngppmktsptv 480
 OY 481 YPVNULPARFTHFGEET 497
 DB 481 ypvndlparfthfgeet 497

RESULT 2
 AAM44160
 ID AAM44160 standard; Protein: 497 AA.
 XX AAM44160;
 AC
 XX
 XX 22-JUN-1998 (first entry)
 DE Human retinoid metabolising protein hp450RA1.
 XX
 XX Retinoid metabolising protein; P450RA1; retinoid oxidase;
 KW retinoid acid; human; inhibitor; antibody; cancer;
 KW actinic keratosis; oral leukoplakia; head tumour; neck tumour;

KW non-small cell lung carcinoma; basal cell carcinoma;
 KW acute promyelocytic leukaemia; skin cancer; acne; psoriasis;
 KW ichthyosis; therapy; diagnosis; screening.
 XX
 XX Homo sapiens.
 OS
 XX MO9749815-A1.
 PN
 XX
 XX 31-DEC-1997.
 PD
 XX
 XX 23-JUN-1997; 97WO-CA00440.
 PF
 XX
 XX 01-OCT-1996; 96US-0724466.
 PR
 XX 21-JUN-1996; 96US-0667546.
 XX
 XX (TOOH) UNITV QUEENS KINGSTON.
 PA
 XX
 XX Beckett BR, Jones G, Petkovich PM, White JA;
 PI
 XX
 XX WPI: 1998-077178/07.
 DR N-PSDB; AAV12204.
 XX
 XX Retinoid metabolising protein - useful to develop products to treat,
 PT e.g. cancer, actinic keratosis, oral leukoplakia, acne, psoriasis or
 PT ichthyosis

Claim 1; Page 54-55; 110pp; English.

CC This protein comprises a novel human retinoid metabolising protein,
 CC designated hp450RA1. Its amino acid sequence was deduced from a
 CC cDNA clone (see AAV12204) isolated from a retinoic acid-treated NT2
 CC cell library. It includes a haem-binding motif characteristic of
 CC cytochrome P450 proteins. hp450RA1 is a retinoid oxidase that has
 CC the ability to hydroxylate retinoic acid at the 4 position of the
 CC beta ionone ring, and is inducible in epithelial cells exposed to
 CC retinoic acid. Zebrafish, human and mouse P450RA1s (see AAM44159-61)
 CC are claimed. They can be expressed in host cells and used to
 CC metabolize retinoic acid in an organism or cell, in drug screening,
 CC and to raise antibodies useful for inhibiting retinoic acid
 CC hydroxylation for the treatment of cancer, actinic keratosis, oral
 CC leukoplakia, secondary tumours of the head and/or neck, non-small
 CC cell lung carcinomas, basal cell carcinomas, acute promyelocytic
 CC leukaemia, skin cancer, and premalignancy associated with actinic
 CC keratosis, acne, psoriasis and/or ichthyosis.
 CC
 XX
 SQ Sequence 497 AA:

Query Match 100.0%; Score 2611; DB 19; Length 497;
 Best Local Similarity 100.0%; Pred. No. 1.5e-253;
 Matches 497; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MGLPILASALCTFVLPILLFLAATKIMDLVCVSGDRSCALPLPPTMGPFPGFTLQM 60
 DB 1 mglpilasalctfvpillflaaklwdlycvsgrdscalp1ppgtmgpfiftglqm 60
 OY 61 VQRRKFLQMKRRKKGFIYKTHLFGRPYRVAGADVRRILLGDRLVSVHMPASVRTL 120
 DB 61 vqrrkflqmkrrkkygfiykthlfgprvrvmgadvrrillgddrlvsvhmpasvrtll 120
 OY 121 GSGCLSNLHDSHKKORRVIMAFSRREALCYVPVITEEVGSSLEQWLSCGERGLLYYPE 180
 DB 121 gsgclsnlhdseshkgrkvimafsrrealcyvpylteevsgslsegwlsgerglllype 180
 OY 181 VKRLMFRIAMRILLCCEPQLAGDSEQOLVAFEMTNLESLPDPVPSGLYRGMKAR 240
 DB 181 vkrlmfriamrillgceplagdsseqqlvafemtrnlslpdpvpsgilyrgmkar 240
 OY 241 NLIHARIEONIRAKICGLRASEAGOGCKDALQLLTEHSMERGEERLDMQALKOSTFELFG 300
 DB 241 nlharieonirakicglraseagcgckdalqlllehswergerldmqalxgstellifg 300
 OY 301 GHETTASATSLITYLGLYPHYLVQKVRBELSKGLICKSNQDNKIDMELLQOLKTYIGCVI 360

```

|||||
Db 301 ghettasaatslitylglyphvlgkvreeksqgllcksnqndkldmetleqkyagcvi 360
QY 361 KETLRINPPVPGGFRVALKTFELNGYQIPKGMNVYISICDTHDVAEIFTNKEEFNDRFS 420
Db 361 ketlrinppvggfrvalktfelngyqipkgwnvysicdthdvadiftnkeefnptrfs 420
QY 421 APHPEDASRFSFIPFGGRLSCVKGKFEPAKILKIFTELARHCDWQLNGPPTMTKTSPTV 480
Db 421 aphpedasrfsfipfggrlscvkgkfeakillkiftvelarhcdwqlngpptmtksptv 480
QY 481 YPYDNLPARTFHGEI 497
Db 481 ypydnlparrfhgei 497

RESULT 3
AAW37735
ID AAW37735 standard; Protein; 497 AA.
XX
AC AAW37735;
XX
DT 07-JUL-1998 (first entry)
XX
DE Cytochrome P450RAI isoform.
XX
KW Retinoid regulated gene; cytochrome P450 gene; enzyme;
KM oxidative metabolism; P450RAI; retinoic acid; RA; promoter; isoform.
XX
OS Synthetic.
XX
PN MO9749832-A2.
XX
PD 31-DEC-1997.
XX
PE 23-JUN-1997; 97WO-CA00488.
XX
PR 01-OCT-1996; 96US-0724466.
XX
PR 21-JUN-1996; 96US-0667546.
XX
PA (TOOH ) UNIV QUEBENS KINGSTON.
XX
PI Petkovich PM.
XX
DR WPI: 1998-077193/07.
XX
DR N-PSDB; AAV09252.
XX
PT Identifying DNA encoding inducible or suppressible cytochrome P450 -
PT useful in cancer chemotherapy and the treatment of acne and
PT psoriasis
PS Disclosure: Pages 59G-59H; 113pp; English.
XX
OS This amino acid sequence is of an isoform of cytochrome zp450RAI.
CC Its expression is dependent on the presence of retinoic acid (RA).
CC The retinoid-regulated genes such as the inducible cytochrome P450RAI
CC gene specifically metabolises a derivative of the RA. The cytochrome
CC P450 gene in general produces enzymes involved in the oxidative
CC metabolism of endogenous and exogenous compounds. The cytochrome P450
CC nucleotide sequence can be used to induce or suppress the expression
CC of its protein. P450RAI is highly induced by RA in cell lines and
CC tissues. This allows for development of a drug screen using promoters
CC and nucleotide sequences to identify drugs which are useful for
CC reducing the catabolism of RA.
XX
SQ Sequence 497 AA;

```

Query Match 93.3%; Score 2437; DB 19; Length 497;
 Best Local Similarity 93.4%; Pred. No. 4.6e-236;
 Matches 464; Conservative 14; Mismatches 19; Indels 0; Gaps 0;

```

QY 1 MGLPALASALCTEVLPLPLFLAAIKLMDLYCSGRDRSCALPLPPGTMGFPFGETLQM 60
Db 1 mglpallasalctvlpplflaalklwdlycsvrsdrscalplppgtmgfpfgetlqm 60
QY 61 VLQRRKFLQMKRRKRYGYTYKTHLEGRPTVRVAGADNVRILLGDRLVSVHPASVPTIL 120
Db 61 vlqrrkflqmkrrkygyfytktlfgreptvrvmgadnvrillgghrlvsvhmpasvrtll 120
QY 121 GSGCLSNLHDSHKKRKRVIMRAFSEALCEVPYITTEVGSSLEQWISGGERGLVYPE 180
Db 121 gsgclsnlhdschkqrvkvmqafsrqaalqcyvvlvaevsccleqwlscgerqlvype 180
QY 181 VKRLMFRIMRILLCCEPOLAGDSDSEOOIVAEFEEMTNLFSLPIDVPFSGLYRGMKAR 240
Db 181 vkrImfrImrIlLlGcePgpgggedeqqlveafeemtrnlfsIpidvpfsqlyrgvkar 240
QY 241 NLIHARIEONIRAKICGLRASEAGGCKDALQLLIHSWEMERGEIDMQALQSSTELLFG 300
Db 241 nlIharIeenIrakIrrlqatepdgckdalqllIehswergerIdmqalKssTeIlIfg 300
QY 301 GHETTASATSLITYLGILYPHYLOKVRREELKSGLLCKSNODNKLDMETLEQKYIGCVI 360
Db 301 ghettasaatslitylglyphvlgkvreeksqgllcksnqndkldmetleqkyigcvi 360
QY 361 KETLRINPPVPGGFRVALKTFELNGYQIPKGMNVYISICDTHDVAEIFTNKEEFNDRFS 420
Db 361 ketlrinppvggfrvalktfelngyqipkgwnvysicdthdvadiftnkeefnptrfi 420
QY 421 APHPEDASRFSFIPFGGRLSCVKGKFEPAKILKIFTELARHCDWQLNGPPTMTKTSPTV 480
Db 421 aphpedasrfsfipfggrlscvkgkfeakillkiftvelarhcdwqlngpptmtksptv 480
QY 481 YPYDNLPARTFHGEI 497
Db 481 ypydnlparrfhgei 497

```

```

RESULT 4
AAW44161
ID AAW44161 standard; Protein; 497 AA.
XX
AC AAW44161;
XX
DT 22-JUN-1998 (first entry)
XX
DE Mouse retinoid metabolising protein mp450RAI.
XX
KW Retinoid metabolising protein; P450RAI; retinoid oxidase;
KW retinoic acid; mouse; inhibitor; antibody; cancer;
KW actinic keratosis; oral leukoplakia; head tumour; neck tumour;
KW non-small cell lung carcinoma; basal cell carcinoma;
KW acute promyelocytic leukemia; skin cancer; acne; psoriasis;
KW ichthyosis; therapy; diagnosis; screening.
XX
OS Mus musculus.
XX
PN MO9749815-A1.
XX
PD 31-DEC-1997.
XX
PE 23-JUN-1997; 97WO-CA00440.
XX
PR 01-OCT-1996; 96US-0724466.
XX
PR 21-JUN-1996; 96US-0667546.
XX
PA (TOOH ) UNIV QUEBENS KINGSTON.
XX
PI Beckelt BR, Jones G, Petkovich PM, White JA;
XX
DR WPI: 1998-077178/07.
XX
DR N-PSDB; AAV12205.
XX
PT Retinoid metabolising protein - useful to develop products to treat,

```

PT e.g. cancer, actinic keratosis, oral leukoplakia, acne, psoriasis or
 PT Ichthyosis

PS Claim 1; Page 65-66; 110pp; English.

CC This protein comprises a novel mouse retinoid metabolising protein,
 CC designated mp450RAI. Its amino acid sequence was deduced from a
 CC cDNA clone (see AAV12205) isolated from a retinoid acid-treated P19
 CC teratocarcinoma library. It includes a haem-binding motif
 CC characteristic of cytochrome P450 proteins. mp450RAI is a retinoid
 CC oxidase that has the ability to hydroxylate retinoic acid at the 4
 CC position of the beta-ionone ring, and is inducible in epithelial
 CC cells exposed to retinoic acid. Zebrafish, human and mouse P450RAIs
 CC (see AAW44159-61) are claimed. They can be expressed in host cells
 CC and used to metabolize retinoic acid in an organism or cell, in drug
 CC screening, and to raise antibodies useful for inhibiting retinoic
 CC acid hydroxylation for the treatment of cancer, actinic keratosis,
 CC oral leukoplakia, secondary tumours of the head and/or neck,
 CC non-small cell lung carcinomas, basal cell carcinomas, acute
 CC promyelocytic leukaemia, skin cancer, and premalignancy associated
 CC with actinic keratosis, acne, psoriasis and/or ichthyosis.

CC Sequence 497 AA:

Query Match 93.3%; Score 2437; DB 19; Length 497;
 Best Local Similarity 93.4%; Pred. No. 4.6e-236;
 Matches 464; Conservative 14; Mismatches 19; Indels 0; Gaps 0;

QY 1 MGIPALLASALCFEVLPLFLFLAATKIMDLKCYSGDRSCALPLPGTMCPPFGETLQ 60
 DB 1 mgipallasalcfvplllflaalklwdlycvsdrscalplpgtmgfpgfgetlqm 60
 QY 61 VLORRKFLOMKRRKYGFYKTHLFGRPYRVMGADNVRRILGDDRLVSVHWPASVETIL 120
 DB 61 vlorrkflomkrrkygfykthlfgprpyrvmgadnvrrilgddrlvsvhwpasvettll 120
 QY 121 GSGCLSNLHDSHKOKRKVIMAFSREALECYVPVTEEVGSSLEQWLSCGERGLLYPE 180
 DB 121 gsgclsnlhdsstkqkkyvmgafsrleaqlcyvllaevsscleqwlscgerglllype 180
 QY 181 VKRIMFRIMRITLGECPQAGDSEQOLVEAFEMTRNLFSLPIDVPFSGILRGKKAR 240
 DB 181 vkrimfrimritlgecpqagdeqolveafemtrnlfsldpdpfsgilrygvkar 240
 QY 241 NIHARIEONIRAKTICGLRASEAGCGCKDALQLLIHSEMERGERLMOALKOSTETLFG 300
 DB 241 niharieoniraktirrlgategdgcckdalqllshswgerlmgalkgstetllfg 300
 QY 301 GHETTASATSLITYLGYLPHVLOKVRRELKSKGLLCKSNODNKLDMEITLQIKYICVYI 360
 DB 301 ghettasatstlitylgyphvlokvrrelkskglcksnodnklmetleqikygcyl 360
 QY 361 KETLRNLNPPVPGFRAVLTPEFNGYQIPKGNVITYSICDTHDVARIENKKEFNDRRS 420
 DB 361 ketlrnlppvpqfvalktfelngyqipkgnvityscdthdvaiftnkeefndrrtl 420
 QY 421 AHPHDASRFSSTIPFGGGRSCVGEKFAKILKIFTELARHCDMOLNPGPTMKRSPVY 480
 DB 421 vphedasarfsstipfggggrscvgekfakillkiftelarhcdmowlngpptomkrspv 480
 QY 481 YPVDNLPAFRTFHGEI 497
 DB 481 ypvdnlpafryfgdi 497

RESULT 5

AAW37733 standard; Protein: 492 AA.

AAW37733;

DT 07-JUL-1998 (first entry)

XX Cytochrome zp450RAI protein.

DE Retinoid regulated gene; cytochrome P450 gene; enzyme;
 KW oxidative metabolism; P450RAI; retinoic acid; RA; promoter.

OS Dario reio.

XX W09749832-A2.

PD 31-DEC-1997.

XX 23-JUN-1997; 97WC-CA00488.

PR 01-OCT-1996; 96US-0724466.

PR 21-JUN-1996; 96US-0667546.

XX (TTOH) UNIV QUEBENS KINGSTON.

PA Petkovich PM.

PI WPI; 1998-077193/07.

DR N-PSDB; AAV09251.

XX Identifying DNA encoding inducible or suppressible cytochrome P450 -
 PT by screening for drugs which reduce the catabolism of retinoic acid,
 PT useful in cancer chemotherapy and the treatment of acne and
 PT psoriasis

XX Example 1; Pages 53-55; 113pp; English.

CC This is the amino acid for cytochrome zp450RAI of the zebra fish.
 CC Its expression is dependent on the presence of retinoic acid (RA).
 CC The retinoid-regulated genes such as the inducible cytochrome P450RAI
 CC gene specifically metabolises a derivative of the RA. The cytochrome
 CC P450 gene in general produces enzymes involved in the oxidative
 CC metabolism of endogenous and exogenous compounds. The cytochrome
 CC P450 nucleotide sequence can be used to induce or suppress the
 CC expression of its protein. P450RAI is highly induced by RA in cell
 CC lines and tissues. This allows for development of a drug screen using
 CC promoters and nucleotide sequences to identify drugs which are useful
 CC for reducing the catabolism of RA.

CC Sequence 492 AA:

Query Match 66.6%; Score 1738; DB 19; Length 492;
 Best Local Similarity 68.0%; Pred. No. 7.5e-166;
 Matches 338; Conservative 63; Mismatches 84; Indels 12; Gaps 5;

QY 1 MGIPALLASALCFEVLPLFLFLAATKIMDLKCYSGDRSCALPLPGTMCPPFGETLQ 60
 DB 1 mgipallmvflcivlpvllflaavaklwmllrrvdpcrplpbgtnlpligetlq 60
 QY 61 VLORRKFLOMKRRKYGFYKTHLFGRPYRVMGADNVRRILGDDRLVSVHWPASVETIL 120
 DB 61 vlorkrkflomkrrkygfykthlfgprpyrvmgadnvrrilgddrlvsvhwpasvettll 120
 QY 121 GSGCLSNLHDSHKOKRKVIMAFSREALECYVPVTEEVGSSLEQWLSCGERGLLYPE 177
 DB 121 gsdclsnlhdsstkqkkyvmgafsrleaqlcyvllaevsscleqwlscgergllly 176
 QY 178 YPEVRKIMFRIMRITLGECP-OLAGDGDSEQOLVEAFEMTRNLFSLPIDVPFSGLYNG 236
 DB 177 ypevmkrlmfrimritlgecpdktl---egelveafemtrnlfsldpdpfsglyng 233
 QY 237 MKARNLIHARIEONIRAKTICGLRASEAGCGCKDALQLLIHSEMERGERLMOALKOSTE 296
 DB 234 ltrarnfinskleenirktktgd-dnenegkykdalqlliensrdsrdsqamkeate 292
 QY 297 LIFGHETTASATSLITYLGYLPHVLOKVRRELKSKGLLCKSNODNKLDMEITLQIKYI 356
 DB 293 lllfgshettasatstlitylgyphvlokvrrelkskglcksnodnklmetlqiklyt 352

PS Claim 1; Page 10; 13pp; Japanese.
XX The present sequence represents a human protein, designated PSEC64, which
CC is related to neuron growth. The PSEC64 protein and its gene can be used
CC for the development of a preventive agent for use in the treatment of
CC diseases in which nerves are involved.
XX
SQ Sequence 216 AA;

Query Match 42.6%; Score 1113; DB 21; Length 216;
Best Local Similarity 99.1%; Pred. No. 1,4e-103;
Matches 214; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGPIALLASALCTFVLPILFLTAIKLMDLYCYSGDRSCALPLPPTGMPFGFTLDM 60
DB 1 mgpllaasalcctfvlplllflaalklmdlycvsgrdscaalp1ppgtmgfpfgetlqm 60
QY 61 VLQRRKFLQMKRRKRYGRIYKTHLFGRTVYVMGADNVRRILLGDDRLVSGVHMFASVRTIL 120
DB 61 vlqrrfqlmqrrrkrygriykythlfgprtvymgadnvrrilllgehlvsvhmpasvrtill 120
QY 121 GSGCLSLNHDSSHKQKKRYMRAFSDEALECYVPTTEYVGSLEQMLSCGERGLLYPE 180
DB 121 gsgclslnhdsshkqkkrymrafsdealecyvpvttteevgsleqmwlsogergllype 180
QY 181 VKRLMFRIMRIILGCEPOLAGDSEQOLVEAFEE 216
DB 181 vkrlmfrimrilllgecpqlagdgdeeqqlveafee 216

RESULT 8
AAG20783 AAG20783 standard; Protein; 468 AA.
XX AAG20783;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 23102.
XX
KM Protein identification: signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PE 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128334.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.

PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134229.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140655.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144095.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.

XX 06-SEP-2000.
PD
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130044.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132487.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 08-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0138847.
PR 16-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.

PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 23-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.

PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.

Query Match	17.3%	Score 450.5	DB 21	Length 461
Best Local Similarity	27.8%	Pred. No. 1	Se-36	
Matches 136	Conservative	96	Mismatches 203	Indels 55
Gaps				15
1	MGPLALASALCTFWPLPLFLAIAIKLMD--LYCVGSGDRSCALPLPGMGPFGEETL	58	:::	
Db	:::		:::	
4	mglllllllvs-lc-----sallrrnqmrlykng-----lppgumyplfigett	44		
QY	59 QWVLQRRKFLQMKRRKRYGFIYTHLFGEPYRVWGAADVNRILLGGDRILVSYHMPASVRI	118		
Db	45 eflkgpnmrnmrglrlysgfsfkshlqgplrlismsevnryllkneekqlyvpyqsgmid	104		
QY	119 ILISGGLSNIHDSHOKRKVKVIMRAFSREALECY-PVITEEVGSSLEOMLSCGERILT-	176		
Db	105 llfctcmnaavhsgshrlmrgslslslsttmrthlhpvkhdfmsyldw---nelevid	161		
QY	177 VVEYEVKRLMFRIAMRILLGCEPOLAGDSDSEQOLVE---AFEEETRNLSLPIDVPSG	232		
Db	162 lqgdktkhmf-----lsltcqia-gnlkpfveefktaffllyvgatlsvpidlpgrtn	212		
QY	233 LYGKMKARNLIHAKIIONIRAKICIGRASGOGCKDLOLLIEHSMREGRRLDMOLKQ	292		
Db	213 yrcqldgrtmtdlltlelner-----rdsgefttdmalylmkkgunrpltd-eetird	265		
QY	293 SSTLELFGHEHTASATSLITLYGLYPHVLQVREELKSGLLCKSNOQKLDMEILEQ	352		
Db	266 qvvellysyvetsttsmmalkylhdpkpaigelaiae-hlafeerkrqdepljledvks	333		
QY	353 LKRTGCIKRTLLNPNPYOGFPAVAKTTELNGYQIDPGMWNIYSICDTHVAELFINKE	412		
Db	324 mkftravlyetsslativngvlrkttrdletjnglylpsgwrllyytreindaaileopl	383		
QY	413 EFNDRSAPHPEDASRFSEIPFGGLRSCVCKEFAKILKILFYTELARHCDWOLLNGPP	472		

```

Db      384  ||| : | : | |||| | ||| : | : | : |
      ifnpwmkksle--sgnscvfvggtrllcpkkelgveissflhytrvryweeigdc 441
QY      473  TMKTSPTVYP 482
      | : |
      442  Im-----VIP 446
Db

```

RESULT 10
AAG46489
ID AAG46489 standard; Protein; 462 AA

AC AAG46489;

DT 18-OCT-2000 (first entry)

AA Arabidopsis thaliana protein fragment SEQ ID NO: 58493.

protein identification; signal transduction pathway; metabolic pathway;

KW termination sequence.

Arabidopsis thaliana

PN EP1033405-A2

PD 06-SEP-2000.

25-FEB-2000; 2000EP-0301439.

AA 25-FEB-1999; 99US-0121825.
PR

PR 09-MAR-1999; 99US-0123548.

PR 25-MAR-1999; 99US-0126264

PR 01-APR-1999; 99US-0127462

PR 08-APR-1999; 99US-0128714

PR 19-APR-1999; 99US-0130077

PR 23-APR-1999; 99US-0130510

PR 28-APR-1999; 99US-0131449

PR 30-APR-1999; 99US-0132407

PR 05-MAY-1999; 99US-0132485

PR 06-MAY-1999; 99US-0132487

PR 11-MAY-1999; 99US-0134256

PR 14-MAY-1999; 99US-0134219

PR 14-MAY-1999; 99US-0134370

PR 19-MAY-1999; 99US-0134941

PR	21-MAY-1999;	99US-0135353
PR	21-MAY-1999;	99US-0135639

PR	25-MAY-1999;	99US-0136021
PD	27-MAY-1999;	99US-0136392

PR	28-MAY-1999;	99US-0136782
DD	01-JUN-1999;	99US-0137222

PR	03-JUN-1999;	99US-0137528
PR	04-JUN-1999;	99US-0137502

PR	07-JUN-1999;	99US-0137724
PR	08-JUN-1999;	99US-0138094

PR	10-JUN-1999;	99US-0138540
PR	10-JUN-1999;	99US-0138847

PR 14-JUN-1999; 9905-0139119

PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144337.
PR 19-JUL-1999; 99US-0144337.
PR 19-JUL-1999; 99US-0144337.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.

PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 25-AUG-1999; 99US-0149930.
PR 26-AUG-1999; 99US-0150566.
PR 27-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151438.
PR 07-SEP-1999; 99US-0151930.
PR 10-SEP-1999; 99US-0152363.
PR 13-SEP-1999; 99US-0153070.
PR 15-SEP-1999; 99US-0153758.
PR 16-SEP-1999; 99US-0154018.
PR 20-SEP-1999; 99US-0154039.
PR 22-SEP-1999; 99US-0154779.
PR 23-SEP-1999; 99US-0155139.
PR 24-SEP-1999; 99US-0155486.
PR 28-SEP-1999; 99US-0155659.
PR 29-SEP-1999; 99US-0156458.
PR 04-OCT-1999; 99US-0156596.
PR 05-OCT-1999; 99US-0157117.
PR 06-OCT-1999; 99US-0157753.
PR 07-OCT-1999; 99US-0157865.
PR 08-OCT-1999; 99US-0158029.
PR 12-OCT-1999; 99US-0158232.
PR 13-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 14-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match

17.3%; Score 450.5; DB 21; Length 462;

Best Local Similarity 27.8%; Pred. No. 1.5e-36;

Matches 136; Conservative 96; Mismatches 203; Indels 55; Gaps 15;

Qy 1 MGLPALASALCTFVLPPLFLFLAIKIMD--LYGVSGDRSCALPLPGTGMGPFPGFTL 58
|||::|||
Db 5 mglllllvs-ic-----sallrtwngmrlytkng-----lppgtmgwplfgelt 45


```
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145291.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 23-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151086.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.

PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 17.38; Score 450.5; DB 21; Length 465;
Best Local Similarity 27.88; Pred. No. 1.5e-36;
Matches 136; Conservative 96; Mismatches 203; Indels 55; Gaps 15;

QY 1 MGDPALASALCTFVLDPLFLAIAIKLMD--LYCVSGRDRSCALPLPGTWGFPFPGTL 58
DB 8 mglillilvs-1c-----sal1rwgmrytkng-----lppgtwmpifigett 48
QY 59 QMVLQRRKFLDMKRRKGYFIYKTHLFGRPVYRVKGNDRIRILIGDRLVSVHWPASVRT 118
DB 49 eflkqgnfmrnqrlrygsffkshllgcplismdsevnryllkneakg1vpyqgsml 108
QY 119 ILGSGCLSNLHDSHKKOKRKIVMRAFSEALCEYV-PVITEEVGSSLEOMLSCGERGL- 176
DB 109 ilgtcmaavhgsghr1mrsglls1sttmrdhllpkvdhfmrsyldq---nelevld 165
QY 177 VPEVKKRMFRIAMRILLGCEPOLAGDGEQOLVE---AFEMFTNLSLPIDVPFSG 232
DB 166 lqdktkhmaf-----lsaltqia--gnlrkpfveefktaffklyvgts1vpidlpqtn 216
QY 233 LYRGMKARNLIHARIEONIRAKICGLRASAGGCKALQULIEHSHREERLDMQALQ 292
DB 217 yrcqilqetnld1l1e1mger-----rdsgetflm1gylmkkegnryl1d-ee1rd 269
QY 293 SSTEILFGHETTASATSLITYGLVPHYLOKREELKSGLLCKNOONKIDMETLEQ 352
DB 270 qvvtllysyetvstsmma1kylhdpkalgelrae--h1atrekrqdeplgievks 327
QY 353 LKYGIVYIKETLRNPVPGGFRVALKTEFLNGYQIPKGNVYISICDHDVAEIFTNKE 412
DB 328 mkttrav1yetsrlat1vngvlrkttrd1e1ngyl1pkyvry1vytre1nydan1yedpl 387
QY 413 EFNPDRFSAPHPEDASRFSIIPFGGIRSCVGKEFAKIIILKIFTVELARHCDQMLNGRP 472
DB 388 ifnpwrmkksle--sqnecfv1g9gtr1cpgk1e1g1ve1ss1lhy1fvtr1ywee1gde 445
QY 473 TMTKSPTVYP 482
DB 446 lm-----vfp 450

RESULT 12
AAC20784 ID AAC20784 standard; Protein; 430 AA.
AC AAC20784;
XX
XX 17-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 23103.
DE
DE Protein identification: signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
XX Arabidopsis thaliana.
OS
```

XX EPI033405-A2.
XX
XX 06-SEP-2000.
PF 25-FEB-2000: 2000EP-0301439.
XX
XX 25-FEB-1999: 99US-0121825.
PR 05-MAR-1999: 99US-0123180.
PR 09-MAR-1999: 99US-0123548.
PR 23-MAR-1999: 99US-0125788.
PR 25-MAR-1999: 99US-0126264.
PR 29-MAR-1999: 99US-0126785.
PR 01-APR-1999: 99US-0127462.
PR 06-APR-1999: 99US-0128234.
PR 16-APR-1999: 99US-0128714.
PR 19-APR-1999: 99US-0129845.
PR 21-APR-1999: 99US-0130077.
PR 23-APR-1999: 99US-0130449.
PR 23-APR-1999: 99US-0130510.
PR 28-APR-1999: 99US-0130891.
PR 30-APR-1999: 99US-0131449.
PR 30-APR-1999: 99US-0132048.
PR 04-MAY-1999: 99US-0132407.
PR 05-MAY-1999: 99US-0132484.
PR 06-MAY-1999: 99US-0132485.
PR 06-MAY-1999: 99US-0132486.
PR 07-MAY-1999: 99US-0132487.
PR 11-MAY-1999: 99US-0132863.
PR 14-MAY-1999: 99US-0134256.
PR 14-MAY-1999: 99US-0134218.
PR 14-MAY-1999: 99US-0134219.
PR 14-MAY-1999: 99US-0134221.
PR 14-MAY-1999: 99US-0134370.
PR 18-MAY-1999: 99US-0134768.
PR 19-MAY-1999: 99US-0134941.
PR 20-MAY-1999: 99US-0135124.
PR 21-MAY-1999: 99US-0135353.
PR 24-MAY-1999: 99US-0135629.
PR 25-MAY-1999: 99US-0136021.
PR 27-MAY-1999: 99US-0136392.
PR 28-MAY-1999: 99US-0136782.
PR 01-JUN-1999: 99US-0137222.
PR 03-JUN-1999: 99US-0137528.
PR 04-JUN-1999: 99US-0137502.
PR 07-JUN-1999: 99US-0137724.
PR 08-JUN-1999: 99US-0138094.
PR 10-JUN-1999: 99US-0138540.
PR 10-JUN-1999: 99US-0138847.
PR 14-JUN-1999: 99US-0139119.
PR 16-JUN-1999: 99US-0139452.
PR 16-JUN-1999: 99US-0139453.
PR 17-JUN-1999: 99US-0139454.
PR 18-JUN-1999: 99US-0139454.
PR 18-JUN-1999: 99US-0139455.
PR 18-JUN-1999: 99US-0139456.
PR 18-JUN-1999: 99US-0139457.
PR 18-JUN-1999: 99US-0139458.
PR 18-JUN-1999: 99US-0139459.
PR 18-JUN-1999: 99US-0139460.
PR 18-JUN-1999: 99US-0139461.
PR 18-JUN-1999: 99US-0139462.
PR 18-JUN-1999: 99US-0139463.
PR 18-JUN-1999: 99US-0139750.
PR 18-JUN-1999: 99US-0139750.
PR 18-JUN-1999: 99US-0139763.
PR 21-JUN-1999: 99US-0139817.
PR 22-JUN-1999: 99US-0139839.
PR 23-JUN-1999: 99US-0140353.
PR 23-JUN-1999: 99US-0140354.
PR 24-JUN-1999: 99US-0140699.
PR 28-JUN-1999: 99US-0140823.
PR 29-JUN-1999: 99US-0140991.
PR 30-JUN-1999: 99US-0141287.
PR 01-JUL-1999: 99US-0141842.
PR 01-JUL-1999: 99US-0142154.
PR 02-JUL-1999: 99US-0142055.
PR 06-JUL-1999: 99US-0142390.
PR 08-JUL-1999: 99US-0142803.
PR 09-JUL-1999: 99US-0142920.
PR 12-JUL-1999: 99US-0142977.
PR 13-JUL-1999: 99US-0143542.
PR 14-JUL-1999: 99US-0143624.
PR 15-JUL-1999: 99US-0144005.
PR 16-JUL-1999: 99US-0144085.
PR 16-JUL-1999: 99US-0144086.
PR 19-JUL-1999: 99US-0144325.
PR 19-JUL-1999: 99US-0144331.
PR 19-JUL-1999: 99US-0144332.
PR 19-JUL-1999: 99US-0144333.
PR 19-JUL-1999: 99US-0144334.
PR 19-JUL-1999: 99US-0144335.
PR 20-JUL-1999: 99US-0144352.
PR 20-JUL-1999: 99US-0144632.
PR 20-JUL-1999: 99US-0144884.
PR 21-JUL-1999: 99US-0144814.
PR 21-JUL-1999: 99US-0145086.
PR 21-JUL-1999: 99US-0145088.
PR 22-JUL-1999: 99US-0145085.
PR 22-JUL-1999: 99US-0145087.
PR 22-JUL-1999: 99US-0145089.
PR 22-JUL-1999: 99US-0145192.
PR 23-JUL-1999: 99US-0145145.
PR 23-JUL-1999: 99US-0145218.
PR 26-JUL-1999: 99US-0145224.
PR 26-JUL-1999: 99US-0145226.
PR 27-JUL-1999: 99US-0145913.
PR 27-JUL-1999: 99US-0145918.
PR 27-JUL-1999: 99US-0145919.
PR 28-JUL-1999: 99US-0145951.
PR 02-AUG-1999: 99US-0146386.
PR 02-AUG-1999: 99US-0146388.
PR 02-AUG-1999: 99US-0146389.
PR 03-AUG-1999: 99US-0147038.
PR 04-AUG-1999: 99US-0147204.
PR 04-AUG-1999: 99US-0147302.
PR 05-AUG-1999: 99US-0147192.
PR 05-AUG-1999: 99US-0147260.
PR 06-AUG-1999: 99US-0147303.
PR 06-AUG-1999: 99US-0147416.
PR 09-AUG-1999: 99US-0147493.
PR 09-AUG-1999: 99US-0147935.
PR 10-AUG-1999: 99US-0148171.
PR 11-AUG-1999: 99US-0148319.
PR 12-AUG-1999: 99US-0148341.
PR 13-AUG-1999: 99US-0148565.
PR 13-AUG-1999: 99US-0148684.
PR 16-AUG-1999: 99US-0149368.
PR 17-AUG-1999: 99US-0149175.
PR 18-AUG-1999: 99US-0149426.
PR 20-AUG-1999: 99US-0149722.
PR 20-AUG-1999: 99US-0149723.
PR 20-AUG-1999: 99US-0149829.
PR 23-AUG-1999: 99US-0149902.
PR 23-AUG-1999: 99US-0149930.
PR 25-AUG-1999: 99US-0150566.
PR 26-AUG-1999: 99US-0150884.
PR 27-AUG-1999: 99US-0151065.
PR 27-AUG-1999: 99US-0151066.
PR 27-AUG-1999: 99US-0151080.
PR 27-AUG-1999: 99US-0151303.
PR 30-AUG-1999: 99US-0151438.
PR 31-AUG-1999: 99US-0151330.
PR 01-SEP-1999: 99US-0152363.
PR 07-SEP-1999: 99US-0153700.
PR 10-SEP-1999: 99US-0153758.
PR 13-SEP-1999: 99US-0154018.

```

PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-015659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159325.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161089.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

Query Match 17.1%; Score 445.5; DB 21; Length 430;
 Best Local Similarity 27.3%; Pred. No. 4.4e-36;
 Matches 124; Conservative 87; Mismatches 201; Indels 43; Gaps 11;

```

QY 49 MCFPEFGTLLQWLR-RFELQKRKRGYFRTKHLFGRTVWGMADNVRILLDDRL 107
DB 1 mwpvylgetlqlysqnrvfftskqkygeifktrlllypcwmlaspeaartivlthahm 60
QY 108 VVHWMPASVRTILSGCLSLNLDSSHOKRKVYIMAFSREALECYVPVITTEVGSLEQM 167
DB 61 fkyprskelkllpsallfhgdyhshlrklvqssfyetetrklkllpdlehlalselqsw 120
QY 168 LSCGERGLLVPEVVKRLMRIMRILLGCEPOLAGDSEQQLVEAFEMEKNLNF----- 222
DB 121 ammpi--vslygemkfkfaavglaifg-----hlesykelkhnylivak 165
QY 223 ---SLPIDVPEGLYRGKMARMLIHARIPONIRAKITGLRASPAGGCKDAQLIENHSM 279
DB 166 gynstfpmalspgsykhalmark-----qlktivselicetretkralq--tdfignhlinfn 219
QY 280 EGEGRILMQALKOSTELLFGSHETTASATSLITYLGLYVHLOKVRBELKS--KGLIC 337
DB 220 ekgyrltgeqialdnllgyfllaqdtascltwllkyl----hddqglleavkaeqkaiye 275
QY 338 KSNONK-LDMILLOLKYICGVIKETRLNPNPVGGFVALKTEPELNYQIIPKGMNVIY 396
DB 276 enstrckkpltrqctnmpklbhkviveslmaslslftfeavdvdeykyllpkygwkmp 335
QY 397 STCDTHDAVEITNKEENPDRFSA-PHPEDASRFSFIPFGGLRSVCYKKEPAKITLKT 455
DB 336 lfrnhmpkytsnpevdprfervnbpkn-----tfmpfgsvnacpgnelaklqilif 390

```

```

QY 456 TVEARHCWQLNGPPTMKTSPVTVNDLPARF 490
DB 391 lhlvsnfwevkgkgykqysfpfpqnglpalf 425

```

RESULT 13

ID AAB07921 standard; protein; 513 AA.

AC AAB07921;

DE 14-NOV-2000 (first entry)

DE A cytochrome P450 enzyme designated DWF4.

KW DWF4; cytochrome P450 enzyme; brassinosteroid; 22alpha-hydroxylation;
 plant phenotype; cell elongation.

OS Arabidopsis sp.

PN W0200047715-A2.

PD 17-AUG-2000.

PE 11-FEB-2000; 2000MO-US03820.

PR 11-FEB-1999; 99US-0119657.

PR 11-FEB-1999; 99US-0119658.

PA (ARIZ-) ARIZONA BOARD OF REGENTS.

PI Azpiroz R, Choe S, Feldmann KA;

DR WPI, 2000-549142/50.

DR N-PSDB; AAA59599.

PT New isolated dwf4 polynucleotide useful for altering the phenotype of
 plants, for diagnostic assays and in the production of antibodies -

PS Claim 50; Fig 11; 113pp; English.

XX The present sequence represents a DWF4 polypeptide. The polypeptide is a
 CC cytochrome P450 enzyme that mediates multiple steps in synthesis of
 CC brassinosteroids. Specifically, it mediates multiple
 CC 22alpha-hydroxylation steps in brassinosteroid biosynthesis. The DWF4
 CC polynucleotide is used for altering the phenotype of a plant. DWF4
 CC plants display a dramatic reduction in the length of different organs,
 CC and this size reduction is attributable to a defect in cell elongation.
 CC The DWF4 polynucleotides and polypeptides can be used in diagnostic
 CC assays and to generate antibodies, which can be used to produce
 CC immunogenic compositions.

XX Sequence 513 AA;

Query Match 16.8%; Score 439; DB 21; Length 513;
 Best Local Similarity 27.0%; Pred. No. 2.5e-35;
 Matches 142; Conservative 85; Mismatches 226; Indels 72; Gaps 15;

```

QY 3 LPALIASALCTFVLPDLFLAIIKMDLVCVSGRDSKALPPLPGTGMGPFPGETL---- 58
DB 10 lpllllpsl-----slllflillk-----rnrktfnlppgasygfllgetlgyik 57
QY 59 -OMVLORRKFLOMKRKRYFYTKHLFGRTVYRWGADNVRILLDDRLVSHWPAVSR 117
DB 58 pytatllgdlmqhyskyklyrslnlfgeptlvsadaglnrflqnegrllfecysprsig 117
QY 118 TTIGSGCLSLNLDSSHOKRKVIMAFSREALECYVPVITTEVGS-----LEQWLSCGR 173
DB 118 gllgkmsmlvlygdhmrhmtsisinfsharlr---tlllkverthlftvldsqg--ns 172
QY 174 GLLVYPEVKRLMRIMRILLGCEPOLAGDSE-QQLVEAFEMEKNLNFSLPIDVPSFG 232

```

Db 173 ifsqdaakkftlnlmakhmnp-----geeteqlkkeyltfmkgvsaplnlpqta 227
Qy 233 LYRGMKARNLTHARIEONIRAKICGLRASEAGOCKDALLIEHSWE-----RGERL 285
Db 228 ykhalgratllkfktekmekrdikfeed-----geeevkteedeemskshvrkqrt 282
Qy 286 DMQ----ALKOS-----LFFGHETTASATSLITLYLGIVPVLOKVRRELK 331
Db 283 dddllgwvlkhsnlisteqldllslflaghesvsvalaalfllgacpkaveelreehl 342
Qy 332 SKGLLCSNDONKMDLMELEOLKYGIVIKETTLNPNVPGGFVAKTFELNGYOIPKG 391
Db 343 eiarakekelgeselnwdykmkmdftgcvinetlrgnvvflnhkalkdvrrykgydilpsg 402
Qy 392 WNVYISICDTVDVAEIFTNKEEFNDRFSAPHEDASRFS-----FPIEGGLRSC 442
Db 403 wkvlpvlsavhldnsrydqpnlflpwrwqgqngasssgsfatwgngnypfgg9p9rlc 462
Qy 443 VGKEPAKILKIFVELARHCDNQLNGPPTMKTSPTVYIPVDNLP 487
Db 463 agselaklemavflhvlvklfnwela-----ddqpfafpvdip 502

RESULT 14

AAG30049 standard; Protein; 481 AA.

AAG30049;

17-OCT-2000 (first entry)

Arabidopsis thaliana protein fragment SEQ ID NO: 35855.

Protein identification; signal transduction pathway; metabolic pathway;
hybridisation assay; genetic mapping; gene expression control; promoter;
termination sequence.

Arabidopsis thaliana.

Ep1033405-A2.

06-SEP-2000.

25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 23-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0138847.
PR 16-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0143542.
PR 13-JUL-1999; 99US-0143624.
PR 14-JUL-1999; 99US-0144005.
PR 15-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 16-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.

PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 10-AUG-1999; 99US-0147935.
PR 11-AUG-1999; 99US-0148317.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 13-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 18-OCT-1999; 99US-0159638.
PR 21-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.

PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 16.6%; Score 434; DB 21; Length 481;
Best Local Similarity 28.0%; Pred. No. 7.4e-35;
Matches 140; Conservative 80; Mismatches 248; Indels 32; Gaps 14;

QY 1 MGLPALASALCTFVLLPLFLAIAIKIMDLXCVSGRDRSCALPIPPGIMGPPFGETIOM 60
Db 1 mwfp111---lg1fv1kwv--lkvnw1yvs1gkckhy--lppgd1gvpvignmsf 52

QY 61 V-----LORRKELOMKRRKYG--FIYKTHLFGRPYVMGADNVRRIILGDRIVSWHP 113
Db 53 lrafktsdpesifgsy1trg1g1ykhnmfgypcv1vtpetctrry1ltdat-h1gwp 111

QY 114 ASVRTIIIGSGCLSLHDSHKKRKKVIMRAFS-REALECYVPVITEVGSSEQLMSGCE 172
Db 112 kstmk1ligrksfvg1sfekhr1rr1tsapvnspealsvy1qflee1vnd1ekws1me 171

QY 173 RGLLVPEVKRIMFRIMRIILGCEPQLAGDSEQLVEAFEMTNTLSLPIDVPSG 232
Db 172 tef1sh--lrlk1tkv1my1f1seseh1vmd----slerey1nl1ygv1amg1n1p1gfa 224

QY 233 LYRGKARNLIHARIEONIRAKITIGLRASEAGGCKDALOLLEHSEWERERLDMAIKQ 292
Db 225 yhr1alkar1kk1vaaf-g1v1tnrrng1kq1issnrk1dml1n1dvk1deng1rv1dde1id 283

QY 293 SSTELLEFGHETASATSLITYLG1XPHVLOKVRRELKSGLLCKSNQDNK1DME1LEQ 352
Db 284 11my1magnessgh1tmwatt1mqenm1l1qakee--der1v1k1rap1qk1il1k1et1re 341

QY 353 LKYIGCVIKETLRLNRPVPGFRYVALKTFELNGYQ1IPKGNV1YS1CDT1DVAE1E1F1NKE 412
Db 342 mvy1sqv1det1rv1t1s1at1freaked1vmd1y1lp1k1ykv1t1w1fnv1n1l1d1p1d1p1d1k 401

QY 413 EFNPDRSAHPEDASFS1IPFGG1RSCVGEKFAK1ILK1ITVELAR1KCDMOL1N-GP 471
Db 402 k1d1ps1rwe1g1t1pk1ag--t1p1p1g1sh1cp1nd1ak1e1s1f1h1n1f1l1k1y1ve1rs1n1p1gc 458

QY 472 PTMKTSEPTVYVNDLPARFT 491
Db 459 pvm-f1p1hnr1pk1nd1ar1c 477

RESULT 15
AAG30048
ID AAG30048 standard; Protein; 489 AA.
XX
AC AAG30048;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 35854.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
hybridisation assay; genetic mapping; gene expression control; promoter;
termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134321.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136382.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139753.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140981.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.

PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144881.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 21-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 09-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 16-SEP-1999; 99US-0154779.
PR 16-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.

```
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161892.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
```

Query Match 16.6%; Score 434; DB 21; Length 489;
Best Local Similarity 28.0%; Pred. No. 7.5e-35;
Matches 140; Conservative 80; Mismatches 248; Indels 32; Gaps 14;

```
QY 1 MGIPALIASALCTFVPLLEFLTAIKIMDLVCVSGDRSCALPLPGTMGPFPGETIQM 60
DB 9 mwfpflll---lgfvlkwv---lkvnwvlyvsklgekkhy---lppgdldgwpvlgmmsf 60
QY 61 V-----LQRRKFLQMKRRKKG--FIYKTHLFGRPTVYMGADNVRRLIGDDRLVSVHWP 113
DB 61 lrafktsdpestiqyitrgtyikamfgypcwlvlttpeccrrvltaddaf-hlgwp 119
QY 114 ASVRTILGSGCLSNLHDSHKQRRKYIMRAFS--REALICYVPVITEEVSGLSEQLSCGE 172
DB 120 kstmkliqkfsfygisfeekhlrlrltsapvnpaalsvyigfifeetvndlekwskmge 179
QY 173 RGLIVPEYKRLMFRIAMRLLCGCEPOLAGDGDSEQDLVEAFEEKTRNLFSLPIDVPSG 232
DB 180 leflsh--lrkltfkvlmvylfssesehvm-----slereynlhyvramglnlpjfa 232
QY 233 LYRGMAKNLIIHARIEONIRAKICGLRASEAGGCKDALQQLIENHSWGERELDMQALQ 292
DB 233 yhnalkarkklyvaaf-qslvtlnrrngkqnlsnrkamlndldvkdengrvlddeellid 291
QY 293 SSTEELFGHEFTTASATSLITYGLYPHVLOKVRBELKSKGLCKSNODNKLDMETLEQ 352
DB 292 lllmylnaghesghltmwacllmgehmlqkakee--gerlyvkkrapgkltlketre 349
QY 353 LKYIGCVIKETRLINPVPVGGFRVALKTEFLNCOYQIPKGNVYISICDTHDVAEIFTNKE 412
DB 350 mwyllsqvldetlrvltfslatreaksdvqmdgylipkgkvtwtfvnhldpeiydpk 409
QY 413 EFNPDERSAPHEIDASRFIFPGGGLRSCVGEFAKILIKITFVELARHCQMLN-GP 471
DB 410 ktdprstwegtytpkag---flpfqigshlcpnndakleisiflhnflklytversnpgc 466
QY 472 PTMKTSTPVYPVDNLPRPFT 491
DB 467 pvm-flphnrpkdnciarit 485
```

Search completed: November 6, 2001, 13:25:25
Job time: 12653 sec


```

QY 61 VLORRKFLOMKRRKYGFIYKTHLFGRPYVMGADVNRILLGDDRLVSVMWPAVSRTIL 120
DB 61 VLORRKFLOMKRRKYGFIYKTHLFGRPYVMGADVNRILLGDDRLVSVMWPAVSRTIL 120
QY 121 GSGCLSNLHDSHKORKKYIMRAFSREALECYVPVITEEVGSSLEQMLSCGERGLLYPE 180
DB 121 GSGCLSNLHDSHKORKKYIMRAFSREALECYVPVITEEVGSSLEQMLSCGERGLLYPE 180
QY 181 VKRLMFIAMRILLGCEPOLAGDSEQOLVEAFEEEMTRNLSLPIDVPPSGLYRGMKAR 240
DB 181 VKRLMFIAMRILLGCEPOLAGDSEQOLVEAFEEEMTRNLSLPIDVPPSGLYRGMKAR 240
QY 241 NLIHARIEONIRAKICGLGRASEAGCGCKDALOLLIEHSMERGERLMOALKOSTELLFG 300
DB 241 NLIHARIEONIRAKICGLGRASEAGCGCKDALOLLIEHSMERGERLMOALKOSTELLFG 300
QY 301 GHETTASAAATSLITYGLYPHYLQKYREELKSKGLCKSNODKLDMEILLEOLKTYGCVI 360
DB 301 GHETTASAAATSLITYGLYPHYLQKYREELKSKGLCKSNODKLDMEILLEOLKTYGCVI 360
QY 361 KETLRINLPVPGGFYALKTFELNGYQIPKGNVYISICDTHDVAEITFNKKEFNDRFS 420
DB 361 KETLRINLPVPGGFYALKTFELNGYQIPKGNVYISICDTHDVAEITFNKKEFNDRFS 420
QY 421 APHPEDASFSFIPIPGGGLRSCVCKEFAKILKIFVELARHCDWOLNGLPPTMTKTSPTV 480
DB 421 APHPEDASFSFIPIPGGGLRSCVCKEFAKILKIFVELARHCDWOLNGLPPTMTKTSPTV 480
QY 481 YPVNMLPARFTHFHEI 497
DB 481 YPVNMLPARFTHFHEI 497

```

RESULT 2

```

; Sequence 4, Application US/08882164D
; Patent No. 6306624
; GENERAL INFORMATION:
; APPLICANT: Petkovich, P. Martin, White, Jay A.
; APPLICANT: Beckett, Barbara R., Jones, Glenville
; TITLE OF INVENTION: Retinoid Metabolizing Protein
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESS: Blake, Cassels & Graydon
; STREET: Box 25, Commerce Court West
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5L 1A9
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 Mb storage
; COMPUTER: COMPAQ, IBM PC compatible
; OPERATING SYSTEM: MS-DOS 5.1
; SOFTWARE: WORD PERFECT
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/882,164D
; FILING DATE: June 25, 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/667,546
; FILING DATE: June 21, 1996
; APPLICATION NUMBER: 08/724,466
; FILING DATE: October 1, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Hunt, John C.
; REGISTRATION NUMBER: 36,424
; REFERENCE/DOCKET NUMBER: 50767/00010
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 863-4344
; TELEFAX: (416) 863-2653
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 497 amino acids
; TYPE: amino acid

```

```

; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-882-164D-4

```

Query Match 100.0%; Score 2611; DB 4; Length 497;
 Best local Similarity 100.0%; Pred. No. 2.3e-279;
 Matches 497; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MGIPALLASALCTFVLPPLLLFLAIFKIMDLVCVSGDRSCALPLPGTMGFPFPGETLQW 60
DB 1 MGIPALLASALCTFVLPPLLLFLAIFKIMDLVCVSGDRSCALPLPGTMGFPFPGETLQW 60
QY 61 VLORRKFLOMKRRKYGFIYKTHLFGRPYVMGADVNRILLGDDRLVSVMWPAVSRTIL 120
DB 61 VLORRKFLOMKRRKYGFIYKTHLFGRPYVMGADVNRILLGDDRLVSVMWPAVSRTIL 120
QY 121 GSGCLSNLHDSHKORKKYIMRAFSREALECYVPVITEEVGSSLEQMLSCGERGLLYPE 180
DB 121 GSGCLSNLHDSHKORKKYIMRAFSREALECYVPVITEEVGSSLEQMLSCGERGLLYPE 180
QY 181 VKRLMFIAMRILLGCEPOLAGDSEQOLVEAFEEEMTRNLSLPIDVPPSGLYRGMKAR 240
DB 181 VKRLMFIAMRILLGCEPOLAGDSEQOLVEAFEEEMTRNLSLPIDVPPSGLYRGMKAR 240
QY 241 NLIHARIEONIRAKICGLGRASEAGCGCKDALOLLIEHSMERGERLMOALKOSTELLFG 300
DB 241 NLIHARIEONIRAKICGLGRASEAGCGCKDALOLLIEHSMERGERLMOALKOSTELLFG 300
QY 301 GHETTASAAATSLITYGLYPHYLQKYREELKSKGLCKSNODKLDMEILLEOLKTYGCVI 360
DB 301 GHETTASAAATSLITYGLYPHYLQKYREELKSKGLCKSNODKLDMEILLEOLKTYGCVI 360
QY 361 KETLRINLPVPGGFYALKTFELNGYQIPKGNVYISICDTHDVAEITFNKKEFNDRFS 420
DB 361 KETLRINLPVPGGFYALKTFELNGYQIPKGNVYISICDTHDVAEITFNKKEFNDRFS 420
QY 421 APHPEDASFSFIPIPGGGLRSCVCKEFAKILKIFVELARHCDWOLNGLPPTMTKTSPTV 480
DB 421 APHPEDASFSFIPIPGGGLRSCVCKEFAKILKIFVELARHCDWOLNGLPPTMTKTSPTV 480
QY 481 YPVNMLPARFTHFHEI 497
DB 481 YPVNMLPARFTHFHEI 497

```

RESULT 3

```

; Sequence 32, Application US/08882164D
; Patent No. 6306624
; GENERAL INFORMATION:
; APPLICANT: Petkovich, P. Martin, White, Jay A.
; APPLICANT: Beckett, Barbara R., Jones, Glenville
; TITLE OF INVENTION: Retinoid Metabolizing Protein
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESS: Blake, Cassels & Graydon
; STREET: Box 25, Commerce Court West
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5L 1A9
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 Mb storage
; COMPUTER: COMPAQ, IBM PC compatible
; OPERATING SYSTEM: MS-DOS 5.1
; SOFTWARE: WORD PERFECT
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/882,164D
; FILING DATE: June 25, 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/667,546
; FILING DATE: June 21, 1996

```

APPLICATION NUMBER: 08/724,466
 FILING DATE: October 1, 1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Hunt, John C.
 REGISTRATION NUMBER: 36,424
 REFERENCE/DOCKET NUMBER: 50767/00010
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (416) 863-4344
 TELEFAX: (416) 863-2653
 INFORMATION FOR SEQ ID NO: 32:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 497 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 US-08-882-164D-32

Query Match 93.3%; Score 2437; DB 4; Length 497;
 Best Local Similarity 93.4%; Pred. No. 3.7e-260;
 Matches 464; Conservative 14; Mismatches 19; Indels 0; Gaps 0;

QY 1 MGIPALLASALCTFVPLPLFLAIAIKLMDLYCVSGDRSCALPLPGTGMGPFPGETLQM 60
 DB 1 MGIPALLASALCTFVPLPLFLAIAIKLMDLYCVSGDRSCALPLPGTGMGPFPGETLQM 60
 QY 61 VLORRKLQMKRRKRYGFIYTHLFGPTVRYMGADNVRRLLDGDDRLVSVHPASVRTL 120
 DB 61 VLORRKLQMKRRKRYGFIYTHLFGPTVRYMGADNVRRLLDGDDRLVSVHPASVRTL 120
 QY 121 GSGCLSNLHDSHOKKRYIMRAFSEALCYVPTVEEYSSLEQMLSCGERGLLYPE 180
 DB 121 GAGCLSNLHDSHOKKRYIMRAFSEALCYVPTVEEYSSLEQMLSCGERGLLYPE 180
 QY 181 VKRLMFIAMRILLCGPOLAGDSEOLVFAFEEMTRNLFLSLPIDVPSGLYRGKAR 240
 DB 181 VKRLMFIAMRILLCGPOLAGDSEOLVFAFEEMTRNLFLSLPIDVPSGLYRGKAR 240
 QY 241 NLHARIEONIRAKICIGLRASEAGCKDALQLLIHSWGERLDMQALKOSTELLFG 300
 DB 241 NLHARIEONIRAKICIGLRASEAGCKDALQLLIHSWGERLDMQALKOSTELLFG 300
 QY 301 GHETTASATSLITYLGLYPHYLVOKVREELKSKGLCKSNODNKLDMETLEQLKICVI 360
 DB 301 GHETTASATSLITYLGLYPHYLVOKVREELKSKGLCKSNODNKLDMETLEQLKICVI 360
 QY 361 KETRLNPNPVGGFRVALKTFELNGYQIPKGMNVIYSICDTHDAVEIFTNKEEFPDRFS 420
 DB 361 KETRLNPNPVGGFRVALKTFELNGYQIPKGMNVIYSICDTHDAVEIFTNKEEFPDRFI 420
 QY 421 APHPEDASRESFIPFGGGLRSCVGEKFAKILKIFTVEIARHCWOLLNGPTMKTSPTV 480
 DB 421 APHPEDASRESFIPFGGGLRSCVGEKFAKILKIFTVEIARHCWOLLNGPTMKTSPTV 480
 QY 481 YPVNDLPAFTFHGEI 497
 DB 481 YPVNDLPAFTFHGEI 497
 RESULT 4
 US-08-724-466B-2
 ; Sequence 2, Application US/08724466B
 ; Patent No. 6063606
 ; GENERAL INFORMATION:
 ; APPLICANT: Petkovich, P. Martin, White, Jay A.,
 ; APPLICANT: Beckett, Barbara R., Jones, Glenville
 ; TITLE OF INVENTION: Retinoid Metabolizing Protein
 ; NUMBER OF SEQUENCES: 30
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESS: Blake, Cassels & Graydon
 ; STREET: Box 25, Commerce Court West
 ; CITY: Toronto
 ; ZIP: M5L 1A9

COUNTRY: Canada
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 Mb storage
 COMPUTER: COMPAQ, IBM PC compatible
 OPERATING SYSTEM: MS-DOS 5.1
 SOFTWARE: WORD PERFECT
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/724,466B
 FILING DATE: October 1, 1996
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/667,546
 FILING DATE: June 21, 1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Hunt, John C.
 REGISTRATION NUMBER: 36,424
 REFERENCE/DOCKET NUMBER: 50767/00004
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (416) 863-4344
 TELEFAX: (416) 863-2653
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 492 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 US-08-724-466B-2

Query Match 66.6%; Score 1738; DB 3; Length 492;
 Best Local Similarity 68.0%; Pred. No. 5.3e-183;
 Matches 338; Conservative 63; Mismatches 84; Indels 12; Gaps 5;

QY 1 MGIPALLASALCTFVPLPLFLAIAIKLMDLYCVSGDRSCALPLPGTGMGPFPGETLQM 60
 DB 1 MGIPALLASALCTFVPLPLFLAIAIKLMDLYCVSGDRSCALPLPGTGMGPFPGETLQM 60
 QY 61 VLORRKLQMKRRKRYGFIYTHLFGPTVRYMGADNVRRLLDGDDRLVSVHPASVRTL 120
 DB 61 VLORRKLQMKRRKRYGFIYTHLFGPTVRYMGADNVRRLLDGDDRLVSVHPASVRTL 120
 QY 121 GSGCLSNLHDSHOKKRYIMRAFSEALCYVPTVEEYSSLEQMLSCGERGLLYPE 180
 DB 121 GSGCLSNLHDSHOKKRYIMRAFSEALCYVPTVEEYSSLEQMLSCGERGLLYPE 180
 QY 178 YPVKRLMFIAMRILLCGPOLAGDSEOLVFAFEEMTRNLFLSLPIDVPSGLYRG 236
 DB 178 YPVKRLMFIAMRILLCGPOLAGDSEOLVFAFEEMTRNLFLSLPIDVPSGLYRG 236
 QY 237 MKARNLIHARIEONIRAKICIGLRASEAGCKDALQLLIHSWGERLDMQALKOSTE 296
 DB 237 MKARNLIHARIEONIRAKICIGLRASEAGCKDALQLLIHSWGERLDMQALKOSTE 296
 QY 297 LIFGGHETTASATSLITYLGLYPHYLVOKVREELKSKGLCKSNODNKLDMETLEQLK 356
 DB 297 LIFGGHETTASATSLITYLGLYPHYLVOKVREELKSKGLCKSNODNKLDMETLEQLK 356
 QY 357 GCVAKETRLNPNPVGGFRVALKTFELNGYQIPKGMNVIYSICDTHDAVEIFTNKEEFP 416
 DB 357 GCVAKETRLNPNPVGGFRVALKTFELNGYQIPKGMNVIYSICDTHDAVEIFTNKEEFP 416
 QY 417 DRESAPHPEDASRESFIPFGGGLRSCVGEKFAKILKIFTVEIARHCWOLLNGPTMK 476
 DB 417 DRESAPHPEDASRESFIPFGGGLRSCVGEKFAKILKIFTVEIARHCWOLLNGPTMK 476
 QY 473 GPTIYPVNDLPAFTFH 489
 DB 473 GPTIYPVNDLPAFTFH 489
 RESULT 5
 US-08-882-164D-2
 ; Sequence 2, Application US/08882164D
 ; Patent No. 6306624


```

;
; REGISTRATION NUMBER: 37,092
; REFERENCE/DOCKET NUMBER: 5051-409
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-854-1400
; TELEFAX: 919-854-1401
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 513 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-948-564-2

```

```

Query Match      10.4%; Score 271.5; DB 3; Length 513;
Best Local Similarity 25.3%; Pred. No. 4,4e-21;
Matches 132; Conservative 71; Mismatches 214; Indels 105; Gaps 23;

```

```

OY 5 ALLASAL-----CTFVLPLLLFLAIAIKLMDLYCVSGDRSCALPLPG 47
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 2 ALLSVLKQLPHELSTHYLVFECIFLILQDI-----RRKKYMLPPS 45
   ||| | | | | | | | | | | | | | | | | | | | | | | | | |
OY 48 TNGFFPFGFTLDM-VLQRRKFLQMKRRKGFYIKTHLFGRPYRVWGADNVRILIGDR 106
   ||| | | | | | | | | | | | | | | | | | | | | | | | | |
DB 46 PPKIPILGMLHQLGTLPHRSFHALS-HKYGPLMLQLGIPLLVSSADVAAREIITKHDV 104
   ||| | | | | | | | | | | | | | | | | | | | | | | | | |
OY 107 IYVSWHPASVFTILSGSCLSNLHDSHKORKKVIYMAFSGREALC-CYVPV----- 155
   ||| | | | | | | | | | | | | | | | | | | | | | | | | |
DB 105 VPSNRQPTAAKIFGCG-----KDAFYIYIEEMRKOKITCKVELMSLKKYRLPH 155
   ||| | | | | | | | | | | | | | | | | | | | | | | | | |
OY 156 -ITTEYVSSLEQWL--SCG-ERGLLVYPEVKRIMFR--IARILIG--CEPOLAGDGS 206
   ||| | | | | | | | | | | | | | | | | | | | | | | | | |
DB 156 SIROEVTELVETALIGACSERCVNLTEM--LMAASNDIYSCVIGRCDDACGSGSS 213
   ||| | | | | | | | | | | | | | | | | | | | | | | | | |
OY 207 EQQIVAEFEEMRNLT-----FSLPIDV-----FSLGYMKARNL-IARIQNIR 252
   ||| | | | | | | | | | | | | | | | | | | | | | | | | |
DB 214 -----SFALDRKIMRLISAFVSGDFPSLGVWVLTSLIPEMKTFLAVDAFDEVIA 267
   ||| | | | | | | | | | | | | | | | | | | | | | | | | |
OY 253 AKICGLIRASEAGQCKDALQLLIEHSWERGE--RLDMQALKOSSTELLFGHETTASAA 309
   ||| | | | | | | | | | | | | | | | | | | | | | | | | |
DB 268 EHESNNKND-----DFLGILQLQ-ECGRIDFQDRDNKLAIVLDMIGSDTTSTL 320
   ||| | | | | | | | | | | | | | | | | | | | | | | | | |
OY 310 TSLIYLYGLYPLVLOKREELKSKGLCKSNODNKLMELIOLKYGIVIKETRLNP 369
   ||| | | | | | | | | | | | | | | | | | | | | | | | | |
DB 321 EMTFAEFLRNPMTKKAQEEVER--VVGINSKAVLDENCVNMNKLKCYVETLRLHP 377
   ||| | | | | | | | | | | | | | | | | | | | | | | | | |
OY 370 VPGGRVALKT---PELNGYOIPKGMNVIYSICDTHVAELTNKEEPDRFSAPHPE- 425
   ||| | | | | | | | | | | | | | | | | | | | | | | | | |
DB 378 LP--LLIARESSSVKLRGYDIPAKTWFINMAIQRDELDMDDEEFLPERFETISQVDL 435
   ||| | | | | | | | | | | | | | | | | | | | | | | | | |
OY 426 DASRFSIFPGGGLRSCVKEFAKILKIFTELARHCDMOL 467
   ||| | | | | | | | | | | | | | | | | | | | | | | | | |
DB 436 NGQDFQLIPFGIGRGCPAMSFGLASTEVYLANLLYWFMMNM 477
   ||| | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

RESULT 13
US-08-313-075A-50
; Sequence 50, Application US/08313075A
; Patent No. 5639870
; GENERAL INFORMATION:
; APPLICANT: Holton, Timothy A.
; APPLICANT: Cornish, Edwin C.
; APPLICANT: Tanaka, Yoshikazu
; TITLE OF INVENTION: GENETIC SEQUENCES ENCODING PLAYONOID
; NUMBER OF SEQUENCES: 58
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Scully, Scott, Murphy & Presser
; STREET: 400 Garden City Plaza
; CITY: Garden City
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 11530

```

```

;
; COMPUTER READABLE FORM:
;
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/313,075A
; FILING DATE: 30-NOV-1994
; CLASSIFICATION: 800
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: AU PL 1538/92
; FILING DATE: 27-MAR-1992
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: AU PL 6698/93
; FILING DATE: 07-JAN-1993
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: AU PCT/AU93/00127
; FILING DATE: 25-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Digilio, Frank S.
; REGISTRATION NUMBER: 31,346
; REFERENCE/DOCKET NUMBER: 9433
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 742-4343
; TELEFAX: (516) 742-4366
; TELEX: 230 901 SANS UR
; INFORMATION FOR SEQ ID NO: 50:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 496 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-313-075A-50

```

```

Query Match      10.1%; Score 264.5; DB 1; Length 496;
Best Local Similarity 24.4%; Pred. No. 2.5e-20;
Matches 126; Conservative 66; Mismatches 205; Indels 119; Gaps 22;

```

```

OY 20 LPLAIAKIMDLICVSG-----RDRSCALPLPFGTMGFPFGFTLQWYQRRKFLQMKRRKY 75
   ||| | | | | | | | | | | | | | | | | | | | | | | | | |
DB 6 ILLGLEFTW--FLVNGMLSLRRRKISKRLPPGFPPLIITGNLLHGNBPKSLAOLAKIH 63
   ||| | | | | | | | | | | | | | | | | | | | | | | | | |
OY 76 GFYIKTHLFGRPYRVWGADNVRILIGD-----RLV-----SVHW-PA--- 114
   ||| | | | | | | | | | | | | | | | | | | | | | | | | |
DB 64 GRIMNKLQGLNTVYISSVAREVLQKODLTFSSNFYDVVHVRHSDFSYVWLPVNSR 123
   ||| | | | | | | | | | | | | | | | | | | | | | | | | |
OY 115 ---SVRTIISGCLIS--NLHDSHKORKKVIYMAFSGREALCYVPVITTEYVSSLEQWLSC 170
   ||| | | | | | | | | | | | | | | | | | | | | | | | | |
DB 124 WTLRKIMNSIFSGNKLDGNOHLRSKV-----QEL---IDYQCKK 162
   ||| | | | | | | | | | | | | | | | | | | | | | | | | |
OY 171 GERGLLVYPEVKRIMFRILMLIGC--EPOLAGD--GDSEQOLVAFEEEMTNL----- 221
   ||| | | | | | | | | | | | | | | | | | | | | | | | | |
DB 163 AKNGEAV--DIGRAFTGTLNLSNTIFSKDLTNPPSDSAKE---FKLVNINVEAGK 216
   ||| | | | | | | | | | | | | | | | | | | | | | | | | |
OY 222 -----FSLPID-----VPSGLYRG-----MKARNLIHARIQNIRAKIC 256
   ||| | | | | | | | | | | | | | | | | | | | | | | | | |
DB 217 PNLVYFPFLKIDIDOGIKRRTNNTFKLGLISGLIDRLKERNL---RDANNT----- 268
   ||| | | | | | | | | | | | | | | | | | | | | | | | | |
OY 257 GLRASEAGQCKDALQLLIEHSWERGEERLDMQALKOSSTELLFGHETTASATSLITYL 316
   ||| | | | | | | | | | | | | | | | | | | | | | | | | |
DB 269 -----DVLDAIINISQENPEEIDRNOIQLCLDLFAAGTDTTSTLEWAMEL 316
   ||| | | | | | | | | | | | | | | | | | | | | | | | | |
OY 317 GLYPLVLOKREELKSKGLCKSNODNKLMELIOLKYGIVIKETRLNPVYGGF-R 375
   ||| | | | | | | | | | | | | | | | | | | | | | | | | |
DB 317 LONPHLOAQOELAO--VIGKGOVEEAD--VGRPLVLRIVAKETLRIHAAPLLPR 371
   ||| | | | | | | | | | | | | | | | | | | | | | | | | |
OY 376 VALKTFELNGYOIPKGMNVIYSICDTHVAELTNKEEPDRFSAPHEDASR--FSPF 434
   ||| | | | | | | | | | | | | | | | | | | | | | | | | |
DB 372 KVEEDVELSTYIIPKDSQYLVVWMAIGRNSDLWENPLVFKPRFWESELDINGRDFELIP 431
   ||| | | | | | | | | | | | | | | | | | | | | | | | | |
OY 435 FGGGLRSCVKEFAKILKIFTELARHCDMOLNG 470
   ||| | | | | | | | | | | | | | | | | | | | | | | | | |

```

Db 432 FGAGRRICPGLPLAMRMPVALGSLNSFNWKLXG 467

RESULT 14

US-09-292-768-68

; Sequence 68, Application US/09292768

; Patent No. 6194185

; GENERAL INFORMATION:

; APPLICANT: Croteau, Rodney B

; APPLICANT: Lupien, Shari L

; APPLICANT: Karp, Frank

; TITLE OF INVENTION: RECOMBINANT MATERIALS AND METHODS FOR THE PRODUCTION OF

; FILE REFERENCE: us013463

; CURRENT APPLICATION NUMBER: US/09/292,768

; CURRENT FILING DATE: 1999-04-14

; EARLIER FILING DATE: 1997-06-24

; NUMBER OF SEQ ID NOS: 70

; SOFTWARE: Patent In Ver. 2.0

; SEQ ID NO 68

; LENGTH: 500

; TYPE: PR1

; ORGANISM: Artificial Sequence

US-09-292-768-68

Query Match 9.7%; Score 252.5; DB 4; Length 500;

Best Local Similarity 21.7%; Pred. No. 5.3e-19;

Matches 114; Conservative 86; Mismatches 217; Indels 109; Gaps 18;

QY 7 LASALCTFVLPLLEFLAIKLMDLYCVSGRSCALPLPPTGTMGPPFGETLQWVLR-- 64

Db 5 ISSAIIILVVTYITSLIIKOW-----RKPKQENLPPGPXKPLIGH-LHLMGKLP 56

QY 65 RKFLQMKRRKRYGFIYKTHL-----FGRPVRYMGADN 96

Db 57 QHALASVAKQYGPVAVHVLQGEVSVLSREATKEAMKLVDPACADRESIGTKIMWYDN 116

QY 97 VRILLGDRL--VSVMPPASVRTILSGCGLSNLHDSHQRKRVIMRAFSEALCEYV 153

Db 117 -----DDIIFSPYSVHM-----RQMRKICVSELLSARVRSFG 149

QY 154 PVITEVGSLSLEQWLSGCEGRLVYPEVKRLMFRIA-MRILLGCEPOLAGDSDSQOLYE 212

Db 150 FIRQDEVSRLLGHLRSSAAAG-----EAVDLTERIATLTCSTICRAAGSVIRHDELY- 203

QY 213 AFEEMTNLFSLPIDVPFSGLYRGKAKRNL-----HARLEQNIIRAKICGLRASE 262

Db 204 ---ELVYDALSMASGFEELADMPSSKLNLLCMNKSILMRMRKRYDALILEAIVEHKLK 260

QY 263 AGQ-GCKDALQLLIEHSMERGERLDM--QALKQSTELLFGGHETTASAAATSLITYGLY 319

Db 261 SGEEFGEDIIDLFRMQKDSQIKVPIITTKAIKAFIDFFSAGTETSSTTTLWMAELMRN 320

QY 320 PHVLQKVEELKSKGLCKSNODNKLMEILBOLYIGCVIKETRLNPPVPGGFVALK 379

Db 321 PEVAKKAQAEVRA--ALKGKTD--WDYDDVQELTYKMSVYKEMRMHPILPLIRSCRE 375

QY 380 TFLNGYOIPKGMVNIYSICDTHVAEIFTNKEENPDRFSAPHEDASR-----FSFI 433

Db 376 ECEVNGYITIPNKARIMINWMSGRNPLTWKEPETFPERF-----DQYSRDEMGNDFEFI 430

QY 434 PFGGGLRSCVCKEPAKILKIFTVELARHCDMQLNG--PPTMKT 477

Db 431 PFGAGRRICPGLNGLANVEVPLAQLLYHFDMKLAEGMNPDSMDMS 476

; APPLICANT: Croteau, Rodney B

; APPLICANT: Lupien, Shari L

; APPLICANT: Karp, Frank

; TITLE OF INVENTION: RECOMBINANT MATERIALS AND METHODS FOR THE PRODUCTION OF

; FILE REFERENCE: us013463

; CURRENT APPLICATION NUMBER: US/09/292,768

; CURRENT FILING DATE: 1999-04-14

; EARLIER FILING DATE: 1997-06-24

; NUMBER OF SEQ ID NOS: 70

; SOFTWARE: Patent In Ver. 2.0

; SEQ ID NO 70

; LENGTH: 500

; TYPE: PR1

; ORGANISM: Artificial Sequence

US-09-292-768-70

Query Match 9.7%; Score 252.5; DB 4; Length 500;

Best Local Similarity 21.7%; Pred. No. 5.3e-19;

Matches 114; Conservative 85; Mismatches 218; Indels 109; Gaps 18;

QY 7 LASALCTFVLPLLEFLAIKLMDLYCVSGRSCALPLPPTGTMGPPFGETLQWVLR-- 64

Db 5 ISSAIIILVVTYITSLIIKOW-----RKPKQENLPPGPXKPLIGH-LHLMGKLP 56

QY 65 RKFLQMKRRKRYGFIYKTHL-----FGRPVRYMGADN 96

Db 57 QHALASVAKQYGPVAVHVLQGEVSVLSREATKEAMKLVDPACADRESIGTKIMWYDN 116

QY 97 VRILLGDRL--VSVMPPASVRTILSGCGLSNLHDSHQRKRVIMRAFSEALCEYV 153

Db 117 -----DDIIFSPYSVHM-----RQMRKICVSELLSARVRSFG 149

QY 154 PVITEVGSLSLEQWLSGCEGRLVYPEVKRLMFRIA-MRILLGCEPOLAGDSDSQOLYE 212

Db 150 FIRQDEVSRLLGHLRSSAAAG-----EAVDLTERIATLTCSTICRAAGSVIRHDELY- 203

QY 213 AFEEMTNLFSLPIDVPFSGLYRGKAKRNL-----HARLEQNIIRAKICGLRASE 262

Db 204 ---ELVYDALSMASGFEELADMPSSKLNLLCMNKSILMRMRKRYDALILEAIVEHKLK 260

QY 263 AGQ-GCKDALQLLIEHSMERGERLDM--QALKQSTELLFGGHETTASAAATSLITYGLY 319

Db 261 SGEEFGEDIIDLFRMQKDSQIKVPIITTKAIKAFIDFFSAGTETSSTTTLWMAELMRN 320

QY 320 PHVLQKVEELKSKGLCKSNODNKLMEILBOLYIGCVIKETRLNPPVPGGFVALK 379

Db 321 PEVAKKAQAEVRA--ALKGKTD--WDYDDVQELTYKMSVYKEMRMHPILPLIRSCRE 375

QY 380 TFLNGYOIPKGMVNIYSICDTHVAEIFTNKEENPDRFSAPHEDASR-----FSFI 433

Db 376 ECEVNGYITIPNKARIMINWMSGRNPLTWKEPETFPERF-----DQYSRDEMGNDFEFI 430

QY 434 PFGGGLRSCVCKEPAKILKIFTVELARHCDMQLNG--PPTMKT 477

Db 431 PFGAGRRICPGLNGLANVEVPLAQLLYHFDMKLAEGMNPDSMDMS 476

Search completed: November 6, 2001, 13:27:05

Job time: 9918 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2001, 13:35:29 ; Search time 71.14 Seconds
(without alignments)
532.173 Million cell updates/sec

Title: US-09-668-482-4

Perfect score: 2611
Sequence: 1 MGIPALLASALCTFVLPLLL.....PTVYVDNLPARFTHNGEI 497

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	663.5	25.4	444	1	cytochrome P450 -
2	519.5	19.9	485	2	probable cytochrom
3	512	19.6	457	2	cytochrome P450 -
4	492.5	18.9	482	2	probable cytochrom
5	463.5	17.8	490	2	hypothetical prote
6	444	17.0	464	2	cytochrome P450 ho
7	439	16.8	513	2	steroid 22-alpha-h
8	434	16.6	489	2	probable cytochrom
9	433.5	16.6	472	1	cytochrome P450 11
10	406	15.5	457	2	cytochrome P450 11
11	400	15.3	492	1	cytochrome P450 DW
12	388	14.9	519	1	cytochrome P450
13	338.5	13.0	476	1	cytochrome P450 Ry
14	333.5	12.8	512	2	probable steroid 2
15	326.5	12.6	496	1	cytochrome P450 4D
16	326.5	12.5	511	1	laurate omega-hydr
17	325	12.4	486	2	lanosterol 14alpha
18	325	12.4	503	2	lanosterol 14alpha
19	324.5	12.4	512	2	hypothetical prote
20	322	12.3	520	2	ferulate 5-hydroxy
21	321	12.3	503	2	sterol 14alpha-dem
22	320	12.3	472	1	probable monooxygen
23	318.5	12.2	511	1	cytochrome P450 4
24	318.5	12.2	520	2	probable cytochrom
25	318	12.2	503	2	hypothetical prote
26	316	12.1	509	2	cytochrome P450 Cy
27	315.5	12.1	519	2	fatty acid omega-h
28	315.5	12.1	591	2	fatty acid omega-h
29	313.5	12.0	488	2	hypothetical prote

30	312.5	12.0	519	2	JX0331	laurate omega-hydr
31	312	11.9	503	2	JC4759	sterol 14-demethyl
32	309.5	11.9	511	1	B34160	cytochrome P450 4A
33	309	11.8	509	1	S68855	lanosterol 14alpha
34	308	11.8	513	2	JC7120	cytochrome P450 en
35	307.5	11.8	511	1	S25707	cytochrome P450 4D
36	305.5	11.7	453	2	T06475	probable obtusifol
37	305.5	11.7	509	1	O4PRT0	laurate omega-hydr
38	305.5	11.7	530	1	A27491	lanosterol 14alpha
39	304.5	11.7	492	2	T14820	obtusifolliol 14-al
40	304.5	11.7	537	2	JC4534	cytochrome P450 4F
41	303	11.6	441	1	C70616	cytochrome P450 Rv
42	301	11.5	528	1	A31854	lanosterol 14alpha
43	300	11.5	487	1	C71417	cytochrome P450 d1
44	300	11.5	511	1	S41598	cytochrome P450 77
45	295	11.3	520	1	A46661	leukotriene B4 ome

ALIGNMENTS

RESULT 1

S75761
cytochrome P450 - Synechocystis sp. (strain PCC 6803)

N:Alternate names: protein slr0574

N:Contains: oxidoreductase (EC 1.-.-.-)

C:Species: Synechocystis sp.

A:Variety: PCC 6803

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000

C:Accession: S75761

R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas

DNA Res. 3, 109-136, 1996

A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocys

S.

A:Reference number: S74322; MUID:97061201

A:Accession: S75761

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-444 <RAN>

A:Cross-references: EMBL:D64003; GB:AB001339; NID:g1001200; PIDN:BA10496.1; PID:g100

F:251-413/Domain: cytochrome P450 homology <P45>

C:Keywords: chromoprotein; heme; iron; metalloprotein; oxidoreductase

F:391/Binding site: heme iron (Cys) (axial ligand) #status predicted

C:Genetics:

A:Gene: cyp

C:Superfamily: Synechocystis cytochrome P450 slr0574; cytochrome P450 homology

Query Match 25.4% Score 663.5; DB 1; Length 444;

Best local similarity 33.7%; Pred. No. 2, 1e-43;

Matches 152; Conservative 87; Mismatches 183; Indels 29; Gaps 8;

QY	41	ALPLPGTGMGFFGGETLQWVLRKRLQMRKRYGFTYKTHLFGRTVRYGADNVRRI	100
DB	10	SLPIPGDFGLPWLGETLNF-LNDGDFGKRRQOFGPIFRKRLGKNVIFISALANRFL	68
QY	101	LLGDRLVSVMPASVFTILSCGSLNHDSSHKOKKVMRAASREALEYVYVIEEV	160
DB	69	FTKEDEFTQATWPLSTRILTPNALATQMGRIHSRRILVQAFPLPTLDYLRKMGIV	128
QY	161	GSSLPEQWLSGGRGLVPEYKRLMFRIMLLGCEPQLAGDDSEQQLVEAFEEWTRN	220
DB	129	QCYLEQWKANF--VIVPQRLKMTFDVAALFPGKEV-----SQNPQLFPWEITYIG	180
QY	221	LESLPIDVPSGLYRGKARNLIHARIEONIRAKICGLRASEAQGCKDALQILIEHSWE	280
DB	161	LESLPIDPPLNTLFGKSQARALLAELEIKIKAR-----QQQPPSEEDAGILLARD	234
QY	281	RGERIDMAKQSSHELLFGGHETTASATSLITYIGLYPHVLQKRELSKGLCKSN	340
DB	235	NNQPLSLPELKDQILLLLFAGHETLSALSSPCILLGSHSDIRERKVEQN-----KLQ	288

QY 341 QDNKIDMEILQKLYIGCVIKETTLRNPVPGFRAVKTFELNGYQIIPKGNVITYSID 400
 Db 289 LSQELTATETIKMPELDVQLEVLRIIPVGGFRELIDQCFQGFHPKPKGLVYQISQ 348
 QY 401 THDAVEITNKNEENPDFR----SAPHEIDASRFSEIFPGGLRSCVGEKFAKILIKITFT 456
 Db 349 THAPEDLTPDEKEPPEFRFTPDGSATNMP---FAHVPEFGGLRCLCEKFAELMKLFA 405
 QY 457 VELARHCDMOLLNGPP-TMKTSPTVYPVDNL 486
 Db 406 TRLIQPDWILLPGONLELVTPSPRPNL 436

RESULT 2

A84859

probable cytochrome P450 (imported) - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 23-Mar-2001

C:Accession: A84859

R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;

M.; Koo, H.; Moffatt, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon, L.

Naus, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.W.; Venter, J.

Nature 402:761-768, 1999

A>Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A:Reference number: A84420; MUID:20083487

A:Accession: A84859

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-485 <SNO>

A:Cross-references: GB:AE002093; NID:94512670; PIDN:AAD21724.1; GSPDB:GN00139

C:Genetics:

A:Gene: At2g42850

A:Map position: 2

C:Superfamily: Synechocystis cytochrome P450 slr0574; cytochrome P450 homology

C:Keywords: heme; iron; metalloprotein

F/432/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 19.9%; Score 519.5; DB 2; Length 485;

Best Local Similarity 28.3%; Pred. No. 2.9e-32;

Matches 132; Conservative 86; Mismatches 201; Indels 47; Gaps 10;

QY 44 LPPGTMGFPFPGETLQVIOARR-----FLQMKRRKGYFTYKTLRGPRTVRYMGADNV 97
 Db 45 LTFEGSLFLEFLCLISORRFGSSKLPPLPGTGMWPYVGTQDYPVFGQSKR 67
 QY 98 RRLILGGDRIVSNWPAVYTIILSGCLSNLHDSHKQKKVIMRAFSREALCEYVIT 157
 Db 105 RLILSNFSLVSSWPSSVOLMGMCIMAKQGRHVLKGIYANSLSYIGLESILPRIC 164
 QY 158 EEVGSLSL-OMLSCGERGLVYPEVKRIAMRIILGCEPOLAGDGSBOQVEAFEE 216
 Db 165 DTVKFNHETEMR--GKEISILYSKAVLFTVVECELYGKIVEIG-----MLEVFER 214
 QY 217 MTRNLISLPIDVFPSCIGRKMKARNLIHARIEONIRAKTIGLGRASEGCKDALDILLIE 276
 Db 215 VLEGVALPVEFPSCSFRARKKARLEITFLVGKREK---RREMEKEAEKENTTLF- 269
 QY 277 HSWERERLDMQALQSSTP-----LLFGHETTASAATSLIYLGLYPRVLOKVR 327
 Db 270 -----SRVLEELIKVITPEEVVDNKNVLLVFAHDTTSTAMSKTFMLQHP---TCR 319
 QY 328 EEIKSGGLCKSN--QDNKIDMEILQKLYIGCVIKETTLRNPVPGFRAVKTFELNG 385
 Db 320 DTLLOHQAQIKRANGGEYLTVEDYKMKMSQOVVRETMRLSPPIGSPFKAAVDIDYGG 379
 QY 386 YQIPKGMNVYISCDTHDAVEITNKNEENPDFRFSAPHEIDASRFSEIFPGGLRSCVGE 445
 Db 380 YTPPKMKMLTMTYGVHYNEIFQDMSFDPTRFDR---IQAYTLPPGGGPRILCAGH 435
 QY 446 EFAKILIKITVELARHCDMOLLNGPP-TMKTSPTVYPVDNLPAFT 491

Db 436 QLAKISILVFMHVVTFGFDWLSLYPDETISMDPLRPPSLGMPKIXIS 481

RESULT 3

T04444

cytochrome P450 - Arabidopsis thaliana

N:Alternate names: protein T18B16.200; protein T5K18.10

C:Contains: oxidoreductase (EC 1.-.-.)

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 28-Jul-2000

C:Accession: T04444; T05806

R:Bayan, M.; Benes, V.; Rechmann, S.; Borkova, D.; Ansoorge, W.; Bancroft, I.; Mewes,

submitted to the Protein Sequence Database, April 1998

A:Reference number: Z15359

A:Accession: T04444

A:Molecule type: DNA

A:Residues: 1-457 <BEV>

A:Cross-references: EMBL:AL021687

A:Experimental source: cultivar Columbia; BAC clone T18B16

R:Bayan, M.; Van Der Schueren, J.; Chhang, Y.J.; Voet, M.; Robben, J.; Volckaert, G.;

submitted to the Protein Sequence Database, April 1998

A:Reference number: Z15453

A:Accession: T05806

A:Molecule type: DNA

A:Residues: 131-457 <BEW>

A:Cross-references: EMBL:AL022580

A:Experimental source: cultivar Columbia; BAC clone T5K18

C:Genetics:

A:Map position: 4

A:Insertions: 67/2; 173/3; 302/3; 358/1; 393/3

A:Note: T18B16.200; T5K18.10

C:Superfamily: Synechocystis cytochrome P450 slr0574; cytochrome P450 homology

C:Keywords: chromoprotein; electron transfer; heme; iron; monooxygenase; oxidoreducta

F/272-433/Domain: cytochrome P450 homology <P45>

Query Match 19.6%; Score 512; DB 2; Length 457;

Best Local Similarity 29.6%; Pred. No. 1e-31;

Matches 144; Conservative 84; Mismatches 204; Indels 54; Gaps 13;

QY 18 LLFLAIAIKMDLYCVSGDR--SCALPLPPGTMGFPFPGETLQVIO-RRKFLQMKRRK 74
 Db 8 LTLFAGSLFLEFLCLISORRFGSSKLPPLPGTGMWPYVGTQDYPVFGQSKR 67
 QY 75 YGFTKTHLREGRPTVRYMGADNVARRILLGDRLVYSVMPASVYTIILSGCLSNLHDSHK 134
 Db 68 YGSVFKTHVLGCPVMTSSPRAAKVLTYSKHLKPTFPASKERMLKQALFFHQGDYHA 127
 QY 135 QRRKVIIRAFSREALCEYVPIYTEEVSLSLEOMLSCGERGLV--YPEVKRIAMRI 192
 Db 128 KLRKLVRALFMPESIRIMWVPDIESIADSLRSW-----EGMINTYQEMKTYTTNVALIS 182
 QY 193 LLGCEPOLAGDGSBOQVEAFEEEMTNLFSLPIDVFPSCIGRKMKARNLIHARIEONIR 252
 Db 183 IFGKDEVL-----YREDLKRCYVILLEGYSMPVNLGTFTHKSKARKREL-----SOIL 232
 QY 253 AKTIGLRASEGCKCKDALDILLIESWGERERLDMQALQSSTELRGGHETTASAATSL 312
 Db 233 ART-----LSERRONGSHNDLLSGFMGDKRELDEQ-IADNIIGVITPAARDTASVMSWI 287
 QY 313 ITYIGLYPHYLQVREELKSKGLCKSNQDNKIDMEILQKLYIGCVIKETTLRNPVPG 372
 Db 288 LKYLAEKNPVLEAVTEE--OMAIRKDKKEGESLTGCTTKMKPLTSRVIOETLRVASILSF 345
 QY 373 GFYRAVKTFELNGYQIIPKGNVITYSIDTHDAVEITNKNEENPDFRFS-APHEIDASRF 431
 Db 346 TFRKAVADVVEYEGLLIPKGVKVLPLFRNIHNSADIFSNNPKFPSPREVAAPKN-----T 400
 QY 432 FIFPGGLRSCVGEKFAKILIKI-----FTVELARHCDMOLLNGPP-TMKTSPTVYPVDNL 487
 Db 401 FMPFGNTHSCPNELAKLEMSIMIHLLTKYKQCVYIYL-----IT 443
 QY 488 ARPTTHF 493

```
Db 444 FSTTHF 449
|||||
RESULT 4
T02739
probable cytochrome P450 At2g29090 [imported] - Arabidopsis thaliana
N:Alternate names: cytochrome P450 homolog T914.17
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 16-Feb-2001
C:Accession: T02739; D84692
R:Rounsley, S.D.; Liu, X.; Kaul, S.; Shea, T.P.; Fujii, C.Y.; Mason, T.M.; Shen, M.; Ron
Submitted to the EMBL Data Library, August 1998
A:Description: Arabidopsis thaliana chromosome II BAC T914 genomic sequence.
A:Reference number: Z14710
A:Accession: T02739
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-482 <STO>
A:Cross-references: EMBL:AC005315; NID:g3461834; PID:g3461849
A:Experimental source: Cultivar Columbia
R:Liu, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
Guss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487
A:Accession: D84692
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-482 <STO>
A:Cross-references: GB:AE002093; NID:g3461849; PID:AA033235.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g29090; T914.17
A:Map position: 2
A:Introns: 80/2; 187/3; 321/3; 351/3; 413/3; 453/2
C:Superfamily: Synechocystis cytochrome P450 slr0574; cytochrome P450 homology
F:291-453/domain: cytochrome P450 homology <P45>

Query Match 18.9%; Score 492.5; DB 2; Length 482;
Best Local Similarity 28.4%; Pred. No. 3.5e-30;
Matches 141; Conservative 90; Mismatches 222; Indels 43; Gaps 12;

4 PALASALCTVPLPLFLAIAIKLMDLYCVSGRDSCLPLPPTGMPFFGETLQWYLG 63
|||||
20 PALITLTIVVVVLLF-----KWLHMKQKQ-----LRPPGSMGLPIGETLRITYTE 68

64 R-RKFLQMKRRKRGYFTYTHLFGRTVYVMGADVNRILLGDDRLVSVHWPASVYRTILGS 122
|||||
69 NNNSFATRONKYGDIFETHILGCPVMISSPEARMYLVSKAHLEKPTYPSPSKERMIGP 128

123 GCLSNHDSHKKORRKYIMRAFSREALCYVPVITEVSSILEQMLSCGERGLLVYPEVK 182
|||||
129 EALFPHQGRYHSTLRKIVOSSFMPSLRPVSHIELLVQLTSSWTSGKSLNTLEY--MK 186

183 RLMFRIAMRILIGCEPQLAGDSDSEQ-----LVAFEEEMRNLSLPIVDPFSGILYGM 237
|||||
187 RYAFYVAI-----MSAFGDKKEEPTTIDVLIKLYORLEKGYSMPLDIPGTLFHKSM 237

236 KARNLIHARIEONIRAKICGRASEAGGCKDALQLLIEHSWGERERLDQALQKOSTEL 297
|||||
238 KKR-----IELSEELKKVILEKRENGREEC--GLLVLLGAKDQKNGISDSOIADNITGV 291

298 LFGHEHTTASATSLITYLGLVPHVLAQV--REELKSKGLLCKSNODNKLDMETILEQLKYI 356
|||||
292 IFAAFTDTASVLTWILKLYLHDPNLLQEVSRQFSIRQIKKENR--RISWEDTRKMLPT 349

357 GCYIKETLLANPPYVGGFVNAKTEFLNQTQPKGMNVIYSICDTHDAEITTNKEEENP 416
|||||
350 TRYIDETLLAASVLSFTFEAAVDVEYDGLLPKGMKVLPLFRRIHNSSEFPDEKFPD 409

417 DFRS-APHPEDASRFSFIIPFGGLRSVCGKEPAKILKIFTEVELARHCDWQLNPGPTMK 475
```

```
Db 410 SRFEVAPKP-----YTYMPGNGVHSCPGSELAKLEMLILHLHTTSFRWEVIGDEGIG 464
|||||
QY 476 TSPYVYPVNDLPARFT 491
|||||
Db 465 YGPFVPYKKKGLPIRVT 480
|||||

RESULT 5
H86185
hypotheical protein [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: H86185
R:theologs, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
Chn, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
ansen, N.F.; Hughes, B.; Hultzer, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Liu, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzia
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719
A:Accession: H86185
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-490 <STO>
A:Cross-references: GB:AE005172; NID:g2388581; PID:AA071462.1; GSPDB:GN00141
C:Genetics:
A:Map position: 1
C:Superfamily: Synechocystis cytochrome P450 slr0574; cytochrome P450 homology

Query Match 17.8%; Score 463.5; DB 2; Length 490;
Best Local Similarity 27.6%; Pred. No. 6.1e-28;
Matches 132; Conservative 85; Mismatches 234; Indels 27; Gaps 11;

25 IKMDLYCVSGRDSCLPLPPTGMPFFGETLQWYLR-----KFLQMKRRKG--F 77
|||||
29 VNWLYSSISGENRHY--LPFGDLGWPFTGNMISFLRAKTSDDPSFTTLTKRGCPKG 85

78 IYKHLEGRPTVYVMGADVNRILLGDDRLVSVHWPASVYRTILGSGCLSNHDSHKKOR 137
|||||
86 IYKHAMFGNPSITVTSDTCRVLTDDAF--KQWPISTWELIGKRSFVGSIEEKKRLR 144

138 KVIMRAFS-REALECYVPVITEVSSILEQMLSCGERGLLVYPEVKRLMIRIARILLGC 196
|||||
145 RLTAAPVNGHEALSTIPIYEENVTYLDKWTMKGEFEFLTH--LRKLTPRIIMYIFLSS 202

197 EPQLAGDSEEQULVEAFEEKTRNLSLPIVDPFSGILRGKARNLIAIEDONIRAKIC 256
|||||
203 ESENVMDA-----DEREYTLANGVRAVAANIPEFAVYHARLAKKTLVAAF--OSITERR 256

257 GLRASEAGGCKDALQLLIEHSWGERERLDQALQKOSTELFGHEHTTASATSLITYL 316
|||||
257 NQRONILSNKKMDLNLNKKDEGKTLDDDELIDVLYLAINAGHSSCHTITMKTAVEL 316

317 GLYPHVLAQVREELKSKGLLCKSNODNKLDMETILE--QLKYICGVIKETLRNLPYPGCF 374
|||||
317 QEHPEVLQRAKAE--QEMILKSRPEGQKGLSKETKRMFELSQVDETLRVITFSLTAF 373

375 RVALKTEFLNQTQPKGMNVIYSICDTHDAEITTNKEEENPDRFSAPHPEDASRFSFI 434
|||||
374 REAATDVEMNVIYLPKGMKVLTPWRVDVHIDEPVPRDKKDPAPAWDGVYKKG--AFLP 431

435 FGGGLRSVCGKEPAKILKIFTEVELARHCDWQLNPGPTKTSPTVYVNDLPARFTH 492
|||||
432 FGAASHLCPGNDLAKLEISIFLHFLKLYQYKRSNPRCPVWYLPHTPRTDNCARISY 489

RESULT 6
```



```

Db 421 VOLWYRSVHMDQVYVDPKFPDSRMEHSPBAG---TFLAFGLARLCPGNDLAKLEIS 477
QY 454 IFT-----VELARRCDWQLNGPPTMKTSPVYVVDNLPARFT 491
Db 478 VELHHFLLGLKLAR-----TMDRCRVRLPAPRPVDNCLAKIT 515

RESULT 13
F70791
cytochrome P450 RV3685c - Mycobacterium tuberculosis (strain H37RV)
N:Contains: oxidoreductase (EC 1.-.-.-)
C:Species: Mycobacterium tuberculosis
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C:Accession: F70791
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:96293587
A:Accession: F70791
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-476 <CO>
A:Cross-references: GB:A1022121; GB:A1123456; NID:g3261559; PIDN:CA18007.1; PID:g296010
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: RV3685c
A:Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology
C:Keywords: Chromoprotein; heme; iron; metalloprotein; oxidoreductase
F:296-444/Domain: cytochrome P450 homology <PA5>
F:422/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 13.0%; Score 338.5; DB 1; Length 476;
Best Local Similarity 25.2%; Pred. No. 2.5e-18;
Matches 102; Conservative 76; Mismatches 178; Indels 49; Gaps 10;

QY 65 KRFLOMKRRKGYFYKTHLFGPRTVAVMGADNVR-----TLGGDRLVSHMPASY 116
Db 68 RMLRLIR- YGIMTPILSLGDAIVSDALAKEVFTAPTDVLLGEGV---GFAA- 121
QY 117 RNLGGCSNLHDSHKKRKVIMAFREALECVVPVTEVSGSLSEQMSCGERGL 176
Db 122 -AIVGSGMFVDEEPEHLRRKLTPRLHGAALDRVPTIENSTRAMATWY--DKPPA 178
QY 177 VYEVKRLMFRITAMRILLCCEPOLAGDSEQOLVAFEEEMTRNLSLPIDVPFS---G 232
Db 179 MLTVARSLMLDVIKVFVD-----DPEVRRLGRPFERLNLGVSEQLTVAYALRLG 233
QY 233 LYRGKARMLIHARIONIRAKTICLRASBAGGCKDALOOLIEHSMGERLDMQALQ 292
Db 234 ALRWPARARANETIDVVMALIAQRADPRLGERHDVLSLVSANGSEQLSDSEIR 293
QY 293 SSPELFGHETTSATSLITYLGLPYHLVKVREELKSGLLCKSNQDNKIDMELQ 352
Db 294 DLITLVLAGHETATATLANAFDILLHHPDALRVRAEDVANGG----- 355
QY 353 LKYGICVIRETLRNPVPGGFVALKTEELANGYQIPKGMNVIYSICDTHDAEIFTNKE 412
Db 336 EATTAIVIMETLKRPRPALTVANAQPLTIGRYEAGRTIVHIIAIRSAYEVHHPH 395
QY 413 EENPDRFSAPHPEDASRFSFIPIGGGLRSCVGEKFAKILIKITV 457
Db 396 EEPREPERFIETRPQ---TYAVPFGGVKRCIGANFS--MKRELITV 435

RESULT 14
H96759
probable steroid 22-alpha-hydroxylase T9L24.44 [Imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001

```

```

C:Accession: H96759
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
ansen, N.E.; Hughes, B.K.; Huizart, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maitl, R.; Marzia
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719
A:Accession: H96759
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-512 <STO>
A:Cross-references: GB:A8005173; NID:g11120803; PIDN:AAG30983.1; GSPDB:GNO0141
C:Genetics:
A:Gene: T9L24.44
A:Map position: 1

Query Match 12.8%; Score 333.5; DB 2; Length 512;
Best Local Similarity 24.9%; Pred. No. 6.6e-18;
Matches 124; Conservative 84; Mismatches 201; Indels 89; Gaps 17;

QY 6 LLSALCTFVLPPLPLAIAKMDLYCVSGDRSCALPFGMGFFPGET----- 57
Db 12 LSVSSSTFFLAFTIIFLLA-----GIARRKRAPRRLPGSGWPLJDTFLAMLANA 64
QY 58 -----LQNVLQRRKFLQMKRKYGYFYKTHLFGPRTVAVMGAD 95
Db 65 GSHPSFVEQIKKFFVSLSCSVLLILKRPDNGSFNIRGRIFSCGLFKRWAVNSADPD 124
QY 96 NVRRILIGDRLVSVHMPASVRTILSGCSNLHDSHKKRKVIMAFREALECT--- 152
Db 125 FNRFIQNEKGLFQSSYKPFRLVGRKGVITVHGDOQRHLSIASMMRHDDQLTFLE 184
QY 153 -VVIYIEVGSLSLEQWLSGGERGLVYPEVKRLMFRITAMRILLCCEPOLAG--DGDSE--Q 209
Db 185 VIVPVMILQTSNFRD---GEVLL-----QDICRKAHLMVN--QLGVSSSESYDE 232
QY 210 LVFAFEEMTRNLSLPIDVPFSGLRGKMARMLIHARIONIRAKICGLASE--AGCGCK 268
Db 233 MSQLESDPFGVCGCLVPIDLGFTYNNKAMKARKELIRKINTIERLQNKASDAGNG-- 290
QY 269 DALQLIEHSMGERLDMQALQKOSTELLFGHETTSATSLITYLGLPYHLQVRE 328
Db 291 -VLGRILLE-----ESLPNESMADFIINLLFAGNETSKTMLFAVYFLTHCPKAMTQLE 344
QY 329 E--LKSGLLCKSNQDNKIDMELQOLKYGICVIRETLRNPVPGG-----FVALKTF 381
Db 345 EHDRLAGML--TWQDYK-----TWDFQCVYIDETLRL-----GGIATVLMREAKEDV 390
QY 382 ELNGYQIPKGMNVIYSICDTHDAEIFTNKEENPDRFSAPHPEDASRFS---FIPFGG 437
Db 391 SYQDYVIRPKGCFVPIPLSAVHLDESYYKESLSFPMWMLDBETQCKRWKRSPTCYCFG 450
QY 438 GLRSCVGEKFAKILIKTF 455
Db 451 GTRCPGAEELARQIALF 468

RESULT 15
S41192
cytochrome P450 4D2 - fruit fly (Drosophila melanogaster)
N:Contains: oxidoreductase (EC 1.-.-.-)
C:Species: Drosophila melanogaster
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C:Accession: S41192; S34291
R:Frolov, M.V.; Alatorsev, V.E.
submitted to the EMBL Data Library, December 1993
A:Description: A cluster of cytochrome P450 genes in the X-chromosome of Drosophila m

```

A;Reference number: S41192
A;Accession: S41192
A;Molecule type: DNA
A;Residues: 1-496 <FR0>
A;Cross-references: EMBL:X/5955; NID:g439650; PIDN:CAA53568.1; PID:g439651
A;Experimental source: strain Oregon R
R;Frollov, M.V.; Alatorsev, V.E.
Submitted to the EMBL Data Library, June 1993
A;Description: Cluster of cytochrome P-450 genes on the X-chromosome in Drosophila melanogaster
A;Reference number: S34291
A;Accession: S34291
A;Molecule type: DNA
A;Residues: 'A', 31-496 <FRM>
A;Cross-references: EMBL:Z23005; NID:g312903; PIDN:CAA80549.1; PID:g312904
A;Experimental source: strain Oregon R
C;Genetics:
A;Gene: Cyp4d2
A;Cross-references: FlyBase:FBgn0011576
A;Map position: X
A;Intons: 53/1; 182/1; 205/2; 392/1
C;Superfamily: human cytochrome P450 CYP4B1; cytochrome P450 homology
C;Keywords: chromoprotein; electron transfer; heme; iron; metalloprotein; oxidoreductase
F;308-471/Domain: cytochrome P450 homology <P45>
F;449/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 12.6% Score 328.5; DB 1; Length 496;
Best Local Similarity 24.7%; Pred. No. 1.5e-17;
Matches 124; Conservative 91; Mismatches 232; Indels 55; Gaps 13;

QY 18 LLLFLAIIKIMDLVCVSGDRSCALPLPPGTMPFFGETLQM---VLQRRKFLQMKR 73
DB 9 LLVAFATLLMDFLW----RRRGILPGRPLPFLGNLAMYRGIDPEQIMDFVKKNQR 63

QY 74 KYGFITYKTHLEGRPVRYMGADNVRRIILGDRLYSVHWPASVRIILSGGCLSNLHDSH 133
DB 64 KYGRLYRWIILHQLAVFSTDPRDIEFLVLSQOHITKNLYKLNCGLDGLMSTGRKWH 123

QY 134 KQKKVIMRAFSREALCEYVPIIEVGSLSQWL--CGERGLVPEVKRMIRIAMR 191
DB 124 -GRKKITPTTHFKILBEVFIPOQSAVMWQLOSRDGMTPINIFVICLTALDITAE 182

QY 192 ILGCE-----POLAGDSEQOLVLEAEKTRN---LFSL--PIDVPSGLYRGM 237
DB 183 TAMGTKINAQKNPMLPYQAVNDVTNLIKRFIHAMQKRVDMIFRLQPTPEAK----RQD 237

QY 238 KARNLIHARIGONIRAKICGL-----RASEAGGCKDALQLLIHSWGERLD 286
DB 238 KAIIVMHDFTENIIRERETLVNNSKETTPREEVNFLOKRRMALLDVLOSTIDGAPLS 297

QY 287 MQALKQSTELLFGHETTASATSLITYLGLYPHVLQVREELKSKGLCKSNQDNKLD 346
DB 298 DEDIREVDVTFMGHDJTTSAISFCLEYISRHPEVQORLOOEIRD--VLGEDRKSPVT 354

QY 347 MEILEQLKYICGVIKETRLMPYPGGRVALKTEFLNGYQIPKQWNIYISICDTHYAE 406
DB 355 LRDIGELKFMENVIKESLRLHPVPMIGRWPAEDVEIRGKHIPAGTNFTMGIFVLLRDE 414

QY 407 IFTNKEEFNPRFASHPEDASRFSFIPFGGLIRSCVCKEFAKILKIFTEVLARHCDWQ 466
DB 415 YFESPDERFRPRFDADYQ-Q-IHPYATIFPSAGPRNCIGOKFAMLEMKSTVSKLRLH--FE 471

QY 467 LLNGPPTMKTS-----PTVY 481
DB 472 LLPLGPEPRHSMNIYCGRPTAF 493

Search completed: November 6, 2001, 13:35:30
Job time: 598 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2001, 13:39:39 ; Search time 62.8 Seconds
(Without alignments)
271.098 Million cell updates/sec

Title: US-09-668-482-4

Perfect score: 1 MGLPALLASALCTEVLPLLL.....PTVYEVNLPARFTFHGEI 497

Sequence: 1 MGLPALLASALCTEVLPLLL.....PTVYEVNLPARFTFHGEI 497

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	2601	99.6	497 1	CP26_HUMAN
2	2437	93.3	497 1	CP26_MOUSE
3	1738	66.6	492 1	CP26_BRARE
4	463.5	17.8	490 1	CP83_ARATH
5	444	17.0	464 1	CP85_LYCES
6	433.5	15.6	472 1	CP91_ARATH
7	400	15.3	492 1	CP36_MYCTU
8	388	14.9	519 1	CP81_MAIZE
9	338.5	13.0	476 1	CP37_MYCTU
10	331	12.7	501 1	CPD2_DROME
11	325.5	12.5	511 1	CP45_RABIT
12	325	12.4	512 1	CP51_RAT
13	324.5	12.3	512 1	CPD1_DROME
14	322	12.3	520 1	CP84_ARATH
15	321	12.3	503 1	CP51_PIG
16	320	12.3	512 1	CPD1_DROME
17	320	12.3	472 1	CP3B_MYCTU
18	318.5	12.2	511 1	CP4C_BLADI
19	317	12.1	511 1	CP3R_ONCMY
20	315.5	12.1	511 1	CP47_RABIT
21	314.5	12.0	493 1	CP132_DROME
22	312.5	12.0	500 1	CP46_MOUSE
23	312.5	12.0	519 1	CP4Y_HUMAN
24	312	11.9	503 1	CP51_HUMAN
25	311.5	11.9	505 1	CP48_DROME
26	308	11.8	507 1	CP3S_BOVIN
27	305.5	11.7	453 1	CP51_WHEAT
28	305.5	11.7	500 1	CP46_HUMAN
29	305.5	11.7	509 1	CP41_RAT
30	305.5	11.7	530 1	CP51_TEAET
31	304.5	11.7	492 1	CP51_SORBI
32	304.5	11.7	537 1	CP56_RAT
33	303.5	11.6	512 1	CP51_CUNNEL

ALIGNMENTS

```
RESULT 1
ID CP26_HUMAN STANDARD; PRT; 497 AA.
AC 043174;
DT 15-DEC-1998 (Rel. 37, last sequence update)
DT 15-DEC-1998 (Rel. 37, last sequence update)
DT 01-OCT-2000 (Rel. 40, last annotation update)
DE CYTOCHROME P450 26 (EC 1.14.-.-) (RETINOIC ACID-METABOLIZING
DE CYTOCHROME) (P450RA1) (RETINOIC ACID 4-HYDROXYLASE).
GN CYP26A1 OR CYP26.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97373542; PubMed=9228017;
RA White J.A., Beckett-Jones B., Guo Y.-D., Dilworth F.J., Bonasoro J.,
RA Jones G., Petkovich M.;
RT "CDNA cloning of human retinoic acid-metabolizing enzyme (hp450RA1)
RT identifies a novel family of cytochromes P450."
RL J. Biol. Chem. 272:18538-18541(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=98380037; PubMed=9716180;
RA Sonneveld E., van den Burg B., van der Leede B.M., Schultes R.K.,
RA Petkovich M., van der Burg B., van der Saag P.T.;
RT "Human retinoic acid (RA) 4-hydroxylase (CYP26) is highly specific for
RT all-trans-RA and can be induced through RA receptors in human breast
RT and colon carcinoma cells."
RL J. Biol. Chem. 272:6329-6337(1998).
RN [3]
RP TISSUE SPECIFICITY.
RX MEDLINE=99045433; PubMed=9826557;
RA Trofinova-Griffin M.E., Juchau M.R.;
RT "Expression of cytochrome P450RA1 (CYP26) in human fetal hepatic and
RT cephalic tissues."
RL Biochem. Biophys. Res. Commun. 252:487-491(1998).
CC -1- FUNCTION: PLAYS A KEY ROLE IN RETINOIC ACID METABOLISM. ACTS ON
CC RETINOIDS, INCLUDING ALL-TRANS-RETINOIC ACID (RA) AND ITS
CC STEREOISOMER 9-CIS-RA. CAPABLE OF BOTH 4-HYDROXYLATION AND 18-
CC HYDROXYLATION. RESPONSIBLE FOR GENERATION OF SEVERAL HYDROXYLATED
CC FORMS OF RA, INCLUDING 4-OH-RA, 4-OXO-RA, AND 18-OH-RA.
CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM.
CC -1- TISSUE SPECIFICITY: HIGHEST LEVELS IN ADULT LIVER, HEART,
CC PITUITARY GLAND, ADRENAL GLAND, PLACENTA AND REGIONS OF THE BRAIN.
CC -1- INDUCTION: BY RETINOIC ACIDS (RA)
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
-----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```

CC -----
DR EMBL: AF005418; AAB88881.1; -
DR MIM: 602239; -
DR InterPro: IPR001128; -
DR Pfam: PF00067; P450; 1.
DR PRINTS: PR00385; P450; 1.
DR PROSITE: PS00086; CYTOCHROME_P450; 1.
DR Oxidoreductase; Monooxygenase; Membrane; Heme; Microsome;
KW Endoplasmic reticulum.
FT BINDING 442 442 HEME (POTENTIAL).
SQ SEQUENCE 497 AA; 56162 MW; EAB6B84B2452EAB3 CRC64;

Query Match
Best local Similarity 99.6%; Score 2601; DB 1; Length 497;
Matches 495; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MGIPALLASALCTFVPLPLFLAIAIKIMDLVCVSGRDSALPLPPTMGPFPGFTLQM 60
DB 1 MGIPALLASALCTFVPLPLFLAIAIKIMDLVCVSGRDSALPLPPTMGPFPGFTLQM 60
QY 61 VLDRRRKFLQKKRRKRYGFIYKTHLFGRTVVMGADNVRRLILGDRLVSVHWPASVETIL 120
DB 61 VLDRRRKFLQKKRRKRYGFIYKTHLFGRTVVMGADNVRRLILGDRLVSVHWPASVETIL 120
QY 121 GSGCLSNLHDSHSHKORRKYIMRAFSEALBCYVPVITEVSSLEQWLSCGERGLVYPE 180
DB 121 GSGCLSNLHDSHSHKORRKYIMRAFSEALBCYVPVITEVSSLEQWLSCGERGLVYPE 180
QY 181 VKRLMPRIARILLGCPROLAGDSDSQOLVEAFEEETRNLFSLPIVPPSGLYRGKAR 240
DB 181 VKRLMPRIARILLGCPROLAGDSDSQOLVEAFEEETRNLFSLPIVPPSGLYRGKAR 240
QY 241 NLHARIEQNIIRAKICGLRASEAGOGCKDALQILIEHSMERGEHLDQALKOSTELLFG 300
DB 241 NLHARIEQNIIRAKICGLRASEAGOGCKDALQILIEHSMERGEHLDQALKOSTELLFG 300
QY 301 GHETTASAATSLITVGLVPHVLQKVEELKSGLLCKSNQDNKLDMEILQKYYICVY 360
DB 301 GHETTASAATSLITVGLVPHVLQKVEELKSGLLCKSNQDNKLDMEILQKYYICVY 360
QY 361 KETLRINLPVPGGFRVALKTFELNGYOIPKGMWVYSICDTHVAELFTKKEPNDRFS 420
DB 361 KETLRINLPVPGGFRVALKTFELNGYOIPKGMWVYSICDTHVAELFTKKEPNDRFS 420
QY 421 APHPEDASRFSFIPIFGGGLSCVGEKFAKILKIFTEVLARHCDMLNPPPTKTSPTV 480
DB 421 LHPEDASRFSFIPIFGGGLSCVGEKFAKILKIFTEVLARHCDMLNPPPTKTSPTV 480
QY 481 YPVNDLPAFTFHFGEI 497
DB 481 YPVNDLPAFTFHFGEI 497

RESULT 2
CP26_MOUSE STANDARD; PRT; 497 AA.
AC 055127;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 40, Last sequence update)
DE CYTOCHROME P450 26 (EC 1.14.-.-) (RETINOIC ACID-METABOLIZING
DE CYTOCHROME) (P450RA1) (RETINOIC ACID 4-HYDROXYLASE).
GN CYP26A1 OR CYP26 OR P450RA.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathu; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C3H;
RX MEDLINE=97392446; PubMed=9250660;
RA Fujii H., Sato T., Kaneko S., Gotoh O., Fujii-Kuriyama Y., Osawa K.,

```

```

RA Kato S., Hamada H.;
RT "Metabolic inactivation of retinoic acid by a novel P450
RT differentially expressed in developing mouse embryos.";
RL EMO J. 16:4163-4173(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=98113212; PubMed=9442090;
RA Abu-Abed S.S., Beckett B.R., Chiba H., Chithalen J.V., Jones G.,
RA Metzger D., Chambon P., Petkovich M.;
RT "Mouse P450RA1 (CYP26) expression and retinoic acid-inducible retinoic
RT acid metabolism in F9 cells are regulated by retinoic acid receptor
RT gamma and retinoid X receptor alpha.";
RL J. Biol. Chem. 273:2409-2415(1998).
CC -1- FUNCTION: PLAYS A KEY ROLE IN RETINOIC ACID METABOLISM. ACTS ON
CC RETINOIDS, INCLUDING ALL-TRANS-RETINOIC ACID (RA) AND ITS
CC STEREOISOMER 9-CIS-RA. CAPABLE OF BOTH 4-HYDROXYLATION AND 18-
CC HYDROXYLATION. RESPONSIBLE FOR GENERATION OF SEVERAL HYDROXYLATED
CC FORMS OF RA, INCLUDING 4-OH-RA, 4-OXO-RA, AND 18-OH-RA.
CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM.
CC -1- INDUCTION: BY RETINOIC ACIDS (RA).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb.slb.ch).
CC -----
DR EMBL: Y12657; CAA73206.1; -
DR MGD; MGI:1096359; Cyp26.
DR InterPro: IPR001128; -
DR Pfam: PF00067; P450; 1.
DR PRINTS: PR00385; P450; 1.
DR PROSITE: PS00086; CYTOCHROME_P450; 1.
KW Oxidoreductase; Monooxygenase; Membrane; Heme; Microsome;
KW Endoplasmic reticulum.
FT BINDING 442 442 HEME (POTENTIAL).
SQ SEQUENCE 497 AA; 56177 MW; 33B07D7C29134471 CRC64;

Query Match
Best local Similarity 93.3%; Score 2437; DB 1; Length 497;
Matches 464; Conservative 14; Mismatches 19; Indels 0; Gaps 0;

QY 1 MGIPALLASALCTFVPLPLFLAIAIKIMDLVCVSGRDSALPLPPTMGPFPGFTLQM 60
DB 1 MGIPALLASALCTFVPLPLFLAIAIKIMDLVCVSGRDSALPLPPTMGPFPGFTLQM 60
QY 61 VLDRRRKFLQKKRRKRYGFIYKTHLFGRTVVMGADNVRRLILGDRLVSVHWPASVETIL 120
DB 61 VLDRRRKFLQKKRRKRYGFIYKTHLFGRTVVMGADNVRRLILGDRLVSVHWPASVETIL 120
QY 121 GSGCLSNLHDSHSHKORRKYIMRAFSEALBCYVPVITEVSSLEQWLSCGERGLVYPE 180
DB 121 GSGCLSNLHDSHSHKORRKYIMRAFSEALBCYVPVITEVSSLEQWLSCGERGLVYPE 180
QY 181 VKRLMPRIARILLGCPROLAGDSDSQOLVEAFEEETRNLFSLPIVPPSGLYRGKAR 240
DB 181 VKRLMPRIARILLGCPROLAGDSDSQOLVEAFEEETRNLFSLPIVPPSGLYRGKAR 240
QY 241 NLHARIEQNIIRAKICGLRASEAGOGCKDALQILIEHSMERGEHLDQALKOSTELLFG 300
DB 241 NLHARIEQNIIRAKICGLRASEAGOGCKDALQILIEHSMERGEHLDQALKOSTELLFG 300
QY 301 GHETTASAATSLITVGLVPHVLQKVEELKSGLLCKSNQDNKLDMEILQKYYICVY 360
DB 301 GHETTASAATSLITVGLVPHVLQKVEELKSGLLCKSNQDNKLDMEILQKYYICVY 360
QY 361 KETLRINLPVPGGFRVALKTFELNGYOIPKGMWVYSICDTHVAELFTKKEPNDRFS 420
DB 361 KETLRINLPVPGGFRVALKTFELNGYOIPKGMWVYSICDTHVAELFTKKEPNDRFS 420

```

```
QY 421 APHEBASRESFIPFGGGLRSCVGEKFAKILKIFVELARHCDMOLLNGPPTKTSPTV 480
DB 421 VPHBEDASRFSEIFPGGGLRSCVGEKFAKILKIFVELARHCDMOLLNGPPTKTSPTV 480
QY 481 YPVNDLPARTFTHGEI 497
DB 481 YPVNDLPARTFTHGEI 497

RESULT 3
CP26_BRARE STANDARD; PRT; 492 AA.
ID CP26_BRARE
AC P79739;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE CYTOCHROME P450 26 (EC 1.14.-.-) (RETINOIC ACID-METABOLIZING
DE CYTOCHROME) (P450RA1) (RETINOIC ACID 4-HYDROXYLASE).
DE CYP26.
GN Brachydanio rerio (zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC Cypriniformes; Cyprinidae; Rasbora; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97094702; PubMed=8939936;
RA Hsu K.E., Dillworth F.J., Jones G., Petkovich M.,
RT "Identification of the retinoic acid-inducible all-trans-retinoic
RT acid 4-hydroxylase."
RL J. Biol. Chem. 271:29922-29927(1996).
CC -1- FUNCTION: PLAYS A KEY ROLE IN RETINOIC ACID METABOLISM. ACTS ON
CC RETINOIDS, INCLUDING ALL-TRANS-RETINOIC ACID (RA) AND ITS
CC STEREOISOMER 9-CIS-RA. CAPABLE OF 4-HYDROXYLATION, HYDROXYLATION,
CC RESPONSIBLE FOR GENERATION OF SEVERAL HYDROXYLATED FORMS OF RA,
CC INCLUDING 4-OH-RA AND 4-OKO-RA.
CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM.
CC -1- INDUCTION: BY RETINOIC ACIDS (RA).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; 068234; AAC60045.1; -
DR ZFIN; ZDB-GENE-990415-44; cyp26.
DR InterPro; IPR001128; -
DR Pfam; PF00067; P450.1.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450.1.
KW Oxidoreductase; Monooxygenase; Membrane; Heme; Microsome;
KW Endoplasmic reticulum.
FT BINDING 438 438 HEME (BY SIMILARITY).
FT SEQUENCE 492 AA; 56281 MW; FD471435B2F30509 CRC64;

Query Match 66.6%; Score 1738; DB 1; Length 492;
Best Local Similarity 68.0%; Pred. No. 3.2e-112;
Matches 338; Conservative 63; Mismatches 84; Indels 12; Gaps 5;

QY 1 MGIPALLASALCTEVPILFLIAIKIMLYCVSGRDSICALPDPGTMGPFPEFTLGM 60
DB 1 MGILYTLMTVFLCTVPLVFLAVALKMLMIRVDFCRSPDPGTMGLPFIGETLQL 60
QY 61 VIQRRKFLQMKRRKGYFIKTLFGFPTVAVGADNVRILLGDRLSVHMPASVFTLL 120
DB 61 IIGRRKFLMKRQKGYCIKTLHFGFPTVAVGADNVRILLGHEKLVSVQMPASVFTLL 120
```

```
QY 121 GSGCLSNLHDSHKKOKKYIMRAFSEALCEVYPTVEEGSSLEBWL---SCGEGTLV 177
DB 121 GSDTLNVHGVOKNNKKALIMRAFSEALCEHYPVIOQEKSAIDQMLDKSC---VLV 176
QY 178 YPEVKRLMFRIMARILLGCEP-OLAGDGSDEQLVEAFEMKRNLFSLPIDVFFSGLYRG 236
DB 177 YPEKRLMFRIMARILLGPEPDKTD---EQELVEAFEMKRNLFSLPIDVFFSGLYRG 233
QY 237 MKARNLIHARIEONIRAKICGLRASPAGCGCDALQLLHESMERKERLDQALQSSSTE 296
DB 234 LRARNFHSKIEENIRKKIÖD-DNENEOYKDALQLLIENSRRSDPEPSLOAMKAAETE 292
QY 297 LIFGHEHTTASATSLITTYGLYPHYLOKVRBEELKSGGLCKSNODNKIDMETLEOLKTY 356
DB 293 LIFGHEHTTASTATSLVMEGLTVEVQKREVOEKVEKMTTPCKGLSMELLDLQKTY 352
QY 357 GCYIKETLRINPVPVGGFRVALKTFELNGYQIPKGNVYISICDTHDVAEIFTKKEFNP 416
DB 353 GCYIKETLRINPVPVGGFRVALKTFELNGYQIPKGNVYISICDTHDVAIVFPNKEEFQ 412
QY 417 DRSAPHEBASRFSEIFPGGGLRSCVGEKFAKILKIFVELARHCDMOLLNGPPTMT 476
DB 413 ERFMSKLEDSRFNYPFGGGLRSCVGEKFAKILKIFVELARHCDMOLLNGPPTMT 472
QY 477 SPYYPVDNLPARTHF 493
DB 473 GPYYPVDNLPTKFTSY 489

RESULT 4
C883_ARATH STANDARD; PRT; 490 AA.
ID C883_ARATH
AC 023051;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE CYTOCHROME P450 88A3 (EC 1.14.-.-).
DE CYP88A3 OR YUP8H12.23.
GN Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=CV. COLUMBIA;
RA Theologis A., Osborne B.I., Vysotskaia V.S., Federspiel N.A.,
RA Toriumi M., Yu G., Oji O., Araujo R., Chung E., Dewar K., Dietrich F.,
RA Ecker J.R., Marzilli A., Oefner P., Davis R.W.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AC000098; AAB71462.1; -
DR InterPro; IPR001128; -
DR Pfam; PF00067; P450.1.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450.1.
KW Oxidoreductase; Monooxygenase; Transmembrane; Heme;
KW TRANSMEM 6 POTENTIAL.
FT BINDING 439 439 HEME (BY SIMILARITY).
FT SEQUENCE 490 AA; 56409 MW; 7FD8CD7A8664D298 CRC64;

Query Match 17.8%; Score 463.5; DB 1; Length 490;
```

[illegible]

FW	Oxidoreductase; Monooxygenase; Membrane; Heme.
FT	BINDING 414 HEME (BY SIMILARITY).
SO	SEQUENCE 464 AA: 53706 MW: D2B21AAAB7B14E94 CRC64:
Query Match	17.0%: Score 444; DB 1; Length 464;
Best Local Similarity	27.4%: Pred. No. 1.7e-23;
Matches 128; Conservative	94; Mismatches 203; Indels 42; Gaps 12;
QY	18 LLLFLAAIKLMDLXCVSGRDRSCALPLPGTGPFFPFSETLQMLVLRKKFLQMKRRKYGF 77
DB	14 LCIFCTALLRMNQVXKNK-----LPPGTGWPLFGFTTEFLLGSPFMKNQARFGS 67
QY	78 IYKTLFGRPRYRWGADNVRKILIGDDRIVSYVHPASVYRTILSGSCISLNHDSHKQK 137
DB	68 FFKSHILCPTIVSDSELNRIYLVNAKGLVPGVQSMIDLKCNIAAVGSAHKYWR 127
QY	138 KVMYAFS-----REALECYVPYTEEVSLSLEOMVSCSGERGLLYPEVKRIFRARIIL 193
DB	128 GALLSLISPTMIRQGL--LPKIDFMSHLTNW--DNKVIDIQEKNNKMAFLSSLQI 181
QY	194 LGCE-PQLAGSDSDSQOLVEAFEEKTRNLFSPLIDVPYSGLYRKKAR---NLIHARIE 248
DB	182 AGIESTSLA-----QEFMSEFFNLVIGTSLIPMLPTNTHRGQARKIYVNLRTLIE 235
QY	249 QNIRAKIGCLRASEAGQCCDALQILIHSMERGEGRLLQMALKOSTSTELFGHEFTTASA 308
DB	236 ER-----RASKRIQ--HMLQGLYLMNEATRFLLTDEMDILIT--ILVSGYETVSTT 284
QY	309 ATSLITTYGLYPRHYQAKREBELKSGLLCKSNQDNKLDELLEQLKTYGCIYKETLRNP 368
DB	285 SMAVKYLIHDPKYLELRKE--HMAIREKKRPEDPIDYNDYRSKRFRAVILLETSLAT 342
QY	369 PVPGCFRALTFTELGNGQIPKGNVIVISCIQDHYVALFETNKKEFNPDRSAPPEDAS 428
DB	343 IVNGVLRTQTQDMELNGIIRKGRVIYTTTELANTDPILYDPYSFNNWRMKDKSLEHON 402
QY	429 RPSFIPGGGLRSCVGKFEFAKILKIFTELARHCDQWLLNGSPPTMK 475
DB	403 --SFLVFGGTGTCROCGKELGVAELSTFHYVTVTKYRMEIGDGDKLMK 447
RESULT 6	
C901ARATH	STANDARD; PRT; 472 AA.
ID C901ARATH	STANDARD; PRT; 472 AA.
AC Q42569	
DT 15-DEC-1998	(Rel. 37, Created)
DT 15-DEC-1998	(Rel. 37, Last sequence update)
DT 15-DEC-1998	(Rel. 37, Last annotation update)
DE CYTOCHROME P450 90A1 (EC 1.14.--).	
CN CYP90A1 OR CYP90 OR CPD	
OS Arabidopsis thaliana (Mouse-ear cress).	
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;	
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;	
OC Brassicales; Brassicaceae; Arabidopsis.	
OX NCBI_Taxid=3702;	
RP SEQUENCE FROM N.A.	
RC STRAIN=CV, COLUMBIA.	
RX MEDLINE=96200769; PubMed=8612270;	
RA Szekeres M., Nemeth K., Koncz-Kalman Z., Mathur J., Kauschmann A.,	
RA Altmann T., Redei G.P., Nagy F., Scheil J., Koncz C.;	
RT "Brassinosteroids rescue the deficiency of CYP90, a cytochrome P450,	
RT controlling cell elongation and de-etiolation in Arabidopsis.",	
RL Cell 85:171-182(1996).	
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.	
CC	
CC This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -	
CC the European Bioinformatics Institute. There are no restrictions on its	
CC use by non-profit institutions as long as its content is in no way	
CC modified and this statement is not removed. Usage by and for commercial	
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/	


```

CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X87367; CAA60793.1; -
CC EMBL; X87368; CAA60794.1; -
CC InterPro; IPR001128; -
CC Pfam; PF00067; P450; 2.
CC PROSITE; PS00086; CYTOCHROME_P450; 1.
CC Oxidoreductase; Monooxygenase; Heme.
CC BINDING 418 418 HEME (BY SIMILARITY).
CC SEQUENCE 472 AA; 53785 MW; 41A73FA6D64E343F CRC64;

Query Match 16.6%; Score 433.5; DB 1; Length 472;
Best Local Similarity 26.5%; Pred. No. 9.3e-23;
Matches 129; Conservative 90; Mismatches 228; Indels 39; Gaps 13;

QY 19 LLEFAAIRKIDLVCSGRDSCALPPTGMPFFEGETLQW-----LQRRKFLQKRR 73
DB 7 LLLLSIAAGFLILLR-RTYRRRNGLPSPSLGLPLIGETPOLGAYKTENDEPIDERVA 65
QY 74 KYGFIYTHLFGRTYVWGANVRILLGDRLVSHVMPASVFTLLGSCCLSLNLDSSH 133
DB 66 RYGSVFMTHLFGEPTLISADPETRRFVLQNEGKLFECSPASICNLGKSHLLMKSLH 125
QY 134 KQKKVIMRAFNRALRCYVY-TTEVSGSLLEQWLSGEGRLVPEVKRIMRI 192
DB 126 KRMHSLTWSFANSSITIDHMLDRLVREFLDSWS-----RVLLMEAKKITFELTVKQ 181
QY 193 LLGCEPOLAGDSEQOLVFAFEEMTRNLFSLPIDVFPSSGLYR-GMKARNLIHARIBONT 251
DB 182 LMSDLP-----GEMSESLRKRYLLVIGFSLPLPL-FSTTYRAIQARR---KVAEAL 231
QY 252 RAKICGLR--ASEAGCGCKDALQLLT--EHSWGERKLDQALKOSTELLFGCHETTAS 307
DB 232 TVVVMKREEREEBEAERKKMLAALLAADGFSDEELVDF-----LVALLVAGYETTTST 285
QY 308 AATSLITVGLGYPHYLQKREELSKGLCKSNQDNKLDMEILLQKYIGVKEITRLN 367
DB 286 IMTLAVKFLTETPLALQKKE--HEKIRAMKSDSYSLSEMDYSMPFTQCVMETLRVA 343
QY 368 PVPDGFVALKTEELNGYQIPKGMNVIYSICDTHDVAEIFTNKEEFNPDRESAPHPEDA 427
DB 344 NLIGVFRRATDVEIKGYKIPKGMKVFSFRAVHLDPNHFKDARTFPMWQNSVTYG 403
QY 428 SRFSPIPPGGGLRSCVGEKEFAKILLKTFYVELARHCDMQLNGPPTKTSPTVYVNDLP 487
DB 404 PSNVFTPPGGGPRCLPGYELARVALSVFLHRLVTGFSW----VPAEDQKLVEFPPTRTQ 458
QY 488 ARFTHF 493
DB 459 KRYPIF 464

RESULT 7
C136 MYCTU STANDARD; PRT; 492 AA.
ID C136 MYCTU
AC P95099;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE PUTATIVE CYTOCHROME P450 136 (EC 1.14.-.-).
GN CYP136 OR RV3059 OR MTCY22D7.22C.
OS Mycobacterium tuberculosis.
OC Bacteria; Filumicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RA MEDLINE=96295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,

```

```

RA Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skellton S., Squares R., Sulston J.E.,
RA Taylor K., Whitehead S., Barrett B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RA complete genome sequence.";
RA Nature 393:537-544 (1998).
CC 1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z83866; CAA06263.1; -
CC Tuberculin; RV3059; -
CC InterPro; IPR001128; -
CC InterPro; IPR002387; -
CC InterPro; IPR002401; -
CC InterPro; IPR002403; -
CC Pfam; PF00067; P450; 1.
CC PRINTS; PR00359; BP450.
CC PRINTS; PR00385; P450.
CC PRINTS; PR00463; EP450I.
CC PRINTS; PR00465; EP450IV.
CC PROSITE; PS00086; CYTOCHROME_P450; 1.
CC Hypothetical protein; Oxidoreductase; Monooxygenase; Heme.
CC BINDING 439 439 HEME (BY SIMILARITY).
CC SEQUENCE 492 AA; 56227 MW; B0A78FCE95622F3D CRC64;

Query Match 15.3%; Score 400; DB 1; Length 492;
Best Local Similarity 28.2%; Pred. No. 1.9e-20;
Matches 137; Conservative 81; Mismatches 192; Indels 76; Gaps 21;

QY 37 DRSCALPRLPPTG-----MGPFPEGELQWLRKRLQMKRRKGYFLYKTH--LFG 85
DB 42 EKRLAEP-PPGSGIKPYVGADGLPLIGHMTEMLRGDPDYLM-----FLYKTKGQVVG 93
QY 86 -----RPTVRWGANVNRILLGDRLVSVH-WPASVFTLLGSCCLSLNLDSSHQKRYV 140
DB 94 DSAVLPGVALGDAQAVIYSNNKDYSSQCGWVPVIGFFHRG-LMLDPEELMFRRLM 152
QY 141 MRAFSEALCEYVPVITEVGSLL-EGWLSGGERGLVPEVKRIMRIAMRILLCEFPQ 199
DB 153 QEAFVRSRLAGYLEQMDRVVSRVADWV-VNDARFLVYPAMKALTLDIASVFWGHEP- 210
QY 200 LAGGDESEQOLVFAFEEMTRN-----LFSLPIDVFPSSGLYRGKARNLH---ARLEN 250
DB 211 -GTDHELVTYVNAKAFITTYTGANAVITVSP---PFL-WMRGLARDELNTFTAYVKE- 264
QY 251 IRAKICGLRASEAGCGCKDALQLLIEHSWGERGLDQALKOSTELLFGCHETTASAT 310
DB 265 -----KREASG---NDLLIVLCQTEDDDGNRFSADLVNMIIFLMAAHDTSTSTAT 313
QY 311 SLITVIGLYPHVLOKREELSKGLCKSNQDNKLDMEILLQKYIGVKEITRLNPPY 370
DB 314 TMAVOLAAHPMQQRCHDESDRHG-----DGPLDIESLEQLESIDLVNNEISIRLVTPV 366
QY 371 PGRFRVALKTEELNGYQIPKGMNVIYSICDTHDVAEIFTNKEEFNPDRESAPHP-DA 429
DB 367 QMANRQTVRDELLGVLPLPGSTNVIAFGMNHRLPELTWDPFLFDERFTEPNEKRRH 426
QY 430 FSFIPFGGLRSCVGEKEFAK-----LKITFYVELARHCDMQLNGPPTKTSPTVY 481
DB 427 YATFPFGGVAKICGMVFDQLEIKTIILHRLRRLRLSLRP-DYQ-----PRMDYSAMP 480
QY 482 PVDNLP 487

```

```
Db 481 PMDMP 486

RESULT 8
C881_MAIZE STANDARD; PRT; 519 AA.
ID C881_MAIZE
AC 043246;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE CYTOCHROME P450 88A1 (EC 1.14.-.-) (DWARF3 PROTEIN).
GN CYP88A1 OR D3.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae;
OC Andropogoneae; Zea.
NCBI_TaxID=4577;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. B73;
RX MEDLINE=96004534; PubMed=7549486;
RA Winkler R.G., Helentjaris T.;
RT "The maize Dwarf3 gene encodes a cytochrome P450-mediated early step
in gibberellin biosynthesis."
RL Plant Cell 7:1307-1317(1995).
CC -1- PATHWAY: EARLY STEP IN GIBBERELLIN BIOSYNTHESIS.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN ROOTS, DEVELOPING LEAVES, THE
VEGETATIVE MERISTEM, AND SUSPENSION CULTURE CELLS.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
CC EMBL: U32579; AAC49067.1; -.
DR InterPro: IPR001128; -.
DR Pfam: PF00067; P450; 1.
DR PROSITE: PS00086; CYTOCHROME_P450; 1.
KW Oxidoreductase; Monooxygenase; Transmembrane; Heme.
FT TRANSMEM 1 21
FT BINDING 466 466 HEME (BY SIMILARITY).
SQ SEQUENCE 519 AA; 57906 MW; 0F8977A024316D95 CRC64;

Query Match 14.9%; Score 388; DB 1; Length 519;
Best Local Similarity 27.2%; Pred. No. 1,4e-19;
Matches 126; Conservative 86; Mismatches 216; Indels 36; Gaps 14;

OY 44 LPPGTMGFPFFGFTLQVLRK-----FLQMKRRKYG--FIYKTHLFGRTVRYVMGADN 96
DB 72 LPPGEGWPLVYGMAFLRFRKSGKPDALFASVYRRGRGVRSFMSPTVLYTTAEG 131
OY 97 VRRILLGDDRLVSVHPASVYRTLLSGCSLNLDSSHQKRVIMRAFSR-EALECVVPV 155
DB 132 CKQVLMDDDAFAFT-GWPKATVALVGRSFVAMPYDEHRRIRKTLAIPINGDALTYLPF 190
OY 156 IPEEVGSSLEQWISGCEGRGLLY-PEVYKRLMRILAMRIILLCERPOLAGDDSDQVLEAF 214
DB 191 IDRTVTSLSLAWMDHG--GSVEFVTELRRTFKIIVQIFG-----GADATYRALERSY 243
OY 215 EEMTNRLFSPLIDVPFSGLYRG-MKARNLTHARIEONIRAKICGLRASFGQCKDALQ 273
DB 244 TELNYGMRAMAILP-GFAYRGALRRARRLYA-VLQGVLDERRAARAKVSGSGVDMMDR 301
OY 274 LIHSNERGERLDMQAKOSSTELLFGHETTASAATSLITYVGLYVHYLVQKREELSK 333
DB 302 LIRAOERGRHLDDEIDIVLVYLNAGHSSGHITMWTATVFLQENDDMPARAKAEQEI 361
OY 334 GLCKSNQDKLDMELLQKLYIGCVIKETLRILNPPVPGGFVALKTFELNGVQIPRGWN 393
```

```
Db 362 MRSIPSSQ-RGLTLDRPKMKVEYLSQVIDETRLVNIISVSPRQATRVVNGYILPRGMK 420
OY 394 VIYSICDTHVALEIFTKKEENPDRESAPHPEDASRSFIPFGGLASCYKGEPAKILK 453
DB 421 VOIWMYRSVHMDPPQVPPPTKFDPSRMEGHSFRAG--TFIAFGAIGALICQNDIAKLEIS 477
OY 454 IFT-----VELARHCDWQLNGPPTKTSPTVYPVNDLPAFT 491
DB 478 VFLHHLFGYKLAR-----TNPGRVRYILPHPRPVNDIAKIT 515

RESULT 9
C137_MYCTU STANDARD; PRT; 476 AA.
ID C137_MYCTU
AC 069653;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE PUTATIVE CYTOCHROME P450 137 (EC 1.14.-.-).
GN CYP137 OR RV3685C OR MTV025.033C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OC NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Taylor S., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E.,
RA Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence."
RL Nature 393:537-544(1998).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
CC EMBL: AL022121; CA118007.1; -.
DR TuberculList; RV3685C; -.
DR InterPro: IPR001128; -.
DR InterPro: IPR002397; -.
DR InterPro: IPR002401; -.
DR Pfam: PF00067; P450; 1.
DR PRINTS: PR00359; BP450.
DR PRINTS: PR00385; P450.
DR PRINTS: PR00463; EP450I.
DR PROSITE: PS00086; CYTOCHROME_P450; 1.
KW Hypothetical protein; Oxidoreductase; Monooxygenase; Heme.
FT BINDING 422 422 HEME (BY SIMILARITY).
SQ SEQUENCE 476 AA; 52265 MW; 4FF0DBCB89BD0548 CRC64;

Query Match 13.0%; Score 338.5; DB 1; Length 476;
Best Local Similarity 25.2%; Pred. No. 3.1e-16;
Matches 102; Conservative 76; Mismatches 178; Indels 49; Gaps 10;

OY 65 RKFLQMKRRRYGFIYKTHLFGRTVRYVMGADNVNR-----ILLGDDRLVSVHPASV 116
DB 68 RRMRLRLRR-YGPIIMPPIILSLDVALIVSDALAKEVFTAPFTVLLGGEV-----GPAA- 121
```


DR FlyBase; FBgn0011576; CYP4d2.
 DR InterPro: IPR001128; -
 DR InterPro: IPR002402; -
 DR Pfam: PF00067; P450; 1.
 DR PRINTS; PRO00385; P450.
 DR PRINTS; PRO0464; EP450I1.
 DR PROSITE; PS00086; CYTOCHROME_P450; 1.
 DR Oxidoreductase; Monooxygenase; Membrane; Heme; Microsome;
 KM Endoplasmic reticulum; Polymorphism.
 FT BINDING 449 449 HEME (BY SIMILARITY).
 FT VARIANT 163 163 K -> M (IN STRAINS CAM-44, CAM-48 AND
 BERKELEY).
 FT CONFLICT 2 25 MISSING (IN REF. 2).
 FT CONFLICT 30 30 I -> A (IN CAA80549).
 FT CONFLICT 160 160 A -> R (IN REF. 1).
 FT CONFLICT 487 501 LRSANGVHGLKRPRA -> CGRTAPFLA (IN REF.
 1).
 SQ SEQUENCE 501 AA; 57914 MW; 8F8D98DBF39FD1A2 CRC64;

Query Match 12.7%; Score 331; DB 1; Length 501;
 Best Local Similarity 24.7%; Pred. No. 1,1e-15;

Matches 120; Conservative 89; Mismatches 231; Indels 46; Gaps 11;

QY 18 LLLFLAIAIKLMDYCVSGRDRSCALPLPGTGMGPFPGFTLQW---VLQRRKFLQMKRR 73
 DB 9 LIAVAFATLLMDLFLM-----RRRGNGILPGRPRLPFLGNLLMWRTGLDPEQIMDFVKKNR 63
 QY 74 KYGFYKTHLEFGPYRVAMADVRRILLDDRLVYHMAVSRTILGSGCLSLHLSSH 133
 DB 64 KYRLRVWTLHQLAVFSTDPRIEFVLSQOHTKNNLYKLNCWIGDGLMSTGRKH 123
 QY 134 KQKKVIMAFSREALECYVPVITEEVGSSLEQWL--CGERGLLYPEVKRLMFRAMR 191
 DB 124 -GRRKITTFFHFKILEQFEYIFDQSAVWVQLOSRADKPTINIPVICTATDIAE 182
 QY 192 ILIGCE-----POLAGDGEQQLVAFEMTRN--LPSL--PIDVPSGLYRGM 237
 DB 183 TANGTINAQKNPNLPVQAIVNDVTNILKRFTHAQVRVMIFELTQPTBAK-----ROD 237
 QY 238 KANLILHARIEQNIIRAKIGL-----RASEAGQCKALQLLIHSWGERLD 286
 DB 238 KAKVMHDFENIIRERETLVNNSKTTPEEEVNLGQRRNALDVLQSTIDGAPLS 297
 QY 287 MQALQSTELTEFGHETPASAATSLITLYGLYVHLQREELSKGLCKSNQDKLD 346
 DB 298 DEDIREVDTEFMEGHDTTSAISFLYEISRHPEVQRIQOEIRD---VLGEDRKSPVT 354
 QY 347 MELLEQLKTYGCIKETRLNPPYPGGFVALKTEFELNGYQIRKGNVYISICDTHVAE 406
 DB 355 LRDGLGELKEMENYIKSLRLHPVPMIGRFAVDVEIRGKHIPAGTMTGIFVLRDPE 414
 QY 407 IFNKKEFNDRSAPRPEDASRSPFIIPGGLRSQVSGKFAKLKTFIVELARHCDWQ 466
 DB 415 YFSPDFEPRERDADVPQ-IHRYAIFPSAGRNICIGQKFAIMEMKSTYSKLIRHELL 473
 QY 467 LINGPP 472
 DB 474 PLGPEP 479

RESULT 11
 CP45_RABIT STANDARD; PRT; 511 AA.
 AC P14579;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE CYTOCHROME_P450 485 PRECURSOR (EC 1.14.15.3) (CYP1A5) (LAURIC ACID
 OMEGA-HYDROXYLASE).
 GN CYP4A5.
 OS Oryctolagus cuniculus (Rabbit).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
 OX NCBI_Taxid=9986;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Kidney;
 RX MEDLINE=90254128; PubMed=2340280;
 RA Johnson E.F., Walker D.L., Griffin K.J., Clark J.E., Okita R.T.,
 RA Neuroff A.S., Masters B.S.;
 RT "Cloning and expression of three rabbit kidney cDNAs encoding lauric
 RT acid omega-hydroxylases";
 RL Biochemistry 29:875-879(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Kidney;
 RX MEDLINE=91192021; PubMed=2013275;
 RA Yokotani N., Kusunose E., Sogawa K., Kawashima H., Kinoshita M.,
 RA Kusunose M., Fujii-Kuriyama Y.;
 RT "cDNA cloning and expression of the mRNA for cytochrome P-450kd which
 RT shows a fatty acid omega-hydroxylating activity";
 RL Eur. J. Biochem. 196:531-536(1991).
 CC -1- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE
 CC MONOOXYGENASES. IN LIVER MICROSOMES, THIS ENZYME IS INVOLVED IN AN
 CC NADPH-DEPENDENT ELECTRON TRANSPORT PATHWAY. IT OXIDIZES A VARIETY
 CC OF STRUCTURALLY UNRELATED COMPOUNDS, INCLUDING STEROIDS, FATTY
 CC ACIDS, AND XENOBIOTICS.
 CC -1- CATALYTIC ACTIVITY: OCTANE + REDUCED RUBREDOXIN + O(2) = 1-OCTANOL
 CC + OXIDIZED RUBREDOXIN + H(2)O.
 CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM.
 CC -1- INDUCTION: P450 CAN BE INDUCED TO HIGH LEVELS IN LIVER AND OTHER
 CC TISSUES BY VARIOUS FOREIGN COMPOUNDS, INCLUDING DRUGS, PESTICIDES,
 CC AND CARCINOGENS.
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; M28655; AAA31229.1; -;
 DR EMBL; X57209; CAA40493.1; -;
 DR PIR; A34260; A34260.
 DR InterPro; IPR001128; -;
 DR InterPro; IPR002402; -;
 DR Pfam; PF00067; P450; 1.
 DR PRINTS; PRO00385; P450.
 DR PRINTS; PRO0464; EP450I1.
 DR PROSITE; PS00086; CYTOCHROME_P450; 1.
 DR Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
 KM Microsome; Endoplasmic reticulum.
 FT PROPEP 1 4 PROBABLE.
 FT CHAIN 5 511 CYTOCHROME_P450 4A5.
 FT BINDING 458 458 HEME (BY SIMILARITY).
 FT CONFLICT 435 435 G -> S (IN REF. 2).
 FT CONFLICT 477 477 V -> L (IN REF. 2).
 SQ SEQUENCE 511 AA; 58357 MW; 11D174BFC8BFA268 CRC64;

Query Match 12.5%; Score 325.5; DB 1; Length 511;
 Best Local Similarity 26.2%; Pred. No. 2.6e-15;
 Matches 137; Conservative 79; Mismatches 221; Indels 85; Gaps 21;

QY 3 LPALASALCTFVL--PLLEFLAIAIKL-----MDLYCVSGRDRSCALPLPGTGMGPFPG 55
 DB 11 LPGLSGLLGVANLGLLLLLKRAQLYLRRQMLRLAQ-----QPPCP----- 55
 QY 56 EPLQMLQRRKFLQMKRRKYGFIYKTHLEFGRPYRVNGADNVR-----RILGDDR 106
 DB 56 -FHWLLGHSREFQMNQLOILKWEKFPACPHWIGKRVQVLYDPDMYKVLIRSD 113
 QY 107 LVSVHMPASVRTILGSGCLNLHDSHKQKKVIMAFSREALECYVPVITEEVGSSLEQ 166

```
Db 114 PERSRSYTFVAVAWIGVGLLT-LNGQPWFOHRRMLLPAPHYDLKRYVGLAMDSVQIMLMDK 172
Oy 167 W-LSCGERGLLVPEVVKRMFRAMRILLGCEPOLAGDSDSEQOLVFAFEEM----- 217
Db 173 WQOLVSDSDLEVFODISIMLTDTIMKCAFVSQGSVDLSRNSQSYIOAVGDLNMLVEAR 232
Oy 218 TNNLSLPLDVFSSLYR-----GKARNLIHARICQNRIRAKICGLRASAGGCK-- 268
Db 233 VNRIFH-----QSDTYRISPEGRLSHRACQLAHEHTORVIOORAAOLO--QEGLEKYR 285
Oy 269 -----DALOLLHESMERGERLDMOALKOSSTELFGHETASATSLITYLGYPHVL 323
Db 286 KRRRLDPLDVLFAKMNKSSSLDDDLRAEVDTFPEGHDTTASVSMFTYLAHHPHQ 345
Oy 324 QKVEBELKSGLLCKSNODNKLMELEOLKYIGCVIKETLLNPPVPGGR--VALKTF 381
Db 346 HRCREI--QGLL---GDGASITWEHLDMQPYTMCIKEMRLYPPVPAISRDLSSPYTF 400
Oy 382 ELNGVQIPKGNVYISICTHVALEFTNKEEFNDRF---SAPHPEDASRSFIPFGG 438
Db 401 P-DGRSLPFGFTVLTSLYGLHNPVWPPEVDFGRFTPGSARSH-----AFLPFSGG 454
Oy 439 LNSCGKEFAKILKTI-FTEELARHCDMQLNGPPTMKTSPT 479
Db 455 ARNCIGKOPAMNELKVAVALTLVR---FELL-PDPRIPIKPT 492

RESULT 12
CP51_RAT
ID CP51_RAT STANDARD: PRT: 503 AA.
AC 064654; 064549;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE CYTOCHROME P450 51 (EC 1.14.14.1) (CYPL1) (P450L1) (STEROL 14-ALPHA
DE DEMETHYLASE) (LANOSTEROL 14-ALPHA DEMETHYLASE) (LDM) (P450-14DM).
GN CYP51.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_Taxid-10116;
RN
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE-94296405; PubMed-8024575;
RA Aoyama Y., Funae Y., Noshiro M., Horiuchi T., Yoshida Y.;
RT "Occurrence of a P450 showing high homology to yeast lanosterol 14-
RT demethylase (P450(14DM)) in the rat liver.";
RL Biochem. Biophys. Res. Commun. 201:1320-1326(1994).
RN
RP SEQUENCE FROM N.A.
RC STRAIN=Wistar;
RX MEDLINE-98158318; PubMed-9498553;
RA Noshiro M., Aoyama Y., Kawamoto T., Gotoh O., Horiuchi T., Yoshida Y.;
RT "Structural and evolutionary studies on sterol 14-demethylase P450
RT (CYP51), the most conserved P450 monooxygenase: I. Structural analyses
RT of the gene and multiple sites of mRNA.";
RL J. Biochem. 122:1114-1121(1997).
RN
RP SEQUENCE OF 18-503 FROM N.A.
RC TISSUE=Liver;
RX MEDLINE-95394364; PubMed-7665087;
RA Sloane D.L., So O.Y., Leung R., Scarafia L.E., Salou N., Jarnagin K.,
RA Swiney D.C.;
RT "Cloning and functional expression of the cDNA encoding rat
RT lanosterol 14-alpha demethylase.";
RL Gene 161:243-248(1995).
CC
CC -1- FUNCTION: CATALYZES C14-DEMETHYLATION OF LANOSTEROL. IT TRANSFORMS
CC LANOSTEROL INTO 4,4'-DIMETHYL CHOLESTA-8,14,24-TRIENE-3-BETA-OL.
CC -1- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) = ROH +
CC OXIDIZED FLAVOPROTEIN + H(2)O.
CC -1- PATHWAY: CHOLESTEROL BIOSYNTHESIS.
```

```
CC -1- SUBCELLULAR LOCATION: MICROSOMAL (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D55681; BAA09529.1; -.
DR EMBL; AB004096; BAA20354.1; JOINED.
DR EMBL; AB004087; BAA20354.1; JOINED.
DR EMBL; AB004088; BAA20354.1; JOINED.
DR EMBL; AB004089; BAA20354.1; JOINED.
DR EMBL; AB004090; BAA20354.1; JOINED.
DR EMBL; AB004091; BAA20354.1; JOINED.
DR EMBL; AB004092; BAA20354.1; JOINED.
DR EMBL; AB004093; BAA20354.1; JOINED.
DR EMBL; AB004094; BAA20354.1; JOINED.
DR EMBL; AB004095; BAA20354.1; JOINED.
DR EMBL; U17697; AAA87074.1; -.
DR InterPro; IPR001128; -.
DR InterPro; IPR002403; -.
DR Pfam; PF00067; P450; 2.
DR PRINTS; PRO0465; EP450TV.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
KW Heme; Cholesterol biosynthesis.
FT BINDING 24 44 POTENTIAL.
FT CONFILCT 181 181 E -> K (IN REF. 3).
SQ SEQUENCE 503 AA; 56706 MW; 33D8F345FE9CF21 CRC64;

Query Match 12.4%; Score 325; DB 1; Length 503;
Best local similarity 22.9%; Pred. No. 2,8e-15;
Matches 117; Conservative 95; Mismatches 236; Indels 62; Gaps 14;

Oy 6 ILASAL--CTFVPLLL-LFLAIKIMLDLYCVSGRDSCLPLPPTGTMG-----PPFG 55
Db 26 LLSTLLIACAFTLSLVYLFRLAV-----GHVQLPAGAKSPYIYSPITPLG 72
Oy 56 ETLQWVADRKFLQMKRRKRYGITYTHLFGRPYRVMGADVNRIL--LGDRLVSWMP 113
Db 73 HAIAFGKSPTEFLNAVRYKGVFSEFTWVGTFYTLISDAAALFNSKNEDLNAAEYTG 132
Oy 114 ASVRTILSGCCLSNHDSHKORKKIVMRASFREALGCVPVITEVGSLEQWLSCGR 173
Db 133 RLTTPVFGKGVAVDPNANVFLQKKIILSGLINAHFKQYVSIIEKEAYEPRKMSGESGR 192
Oy 174 GLVYVPEVKRMFRAMRILLGCEPOLAGDSDSEQOLVFAFEEMTRNL--FS-----LP 225
Db 193 N--VEFALSELITLFAHCHLHGKE-----IRSQLNEKAOLYADLDGGSFHAAMLPL 242
Oy 226 IDVPSSGLRGKMA-----NLIHARIEONIRAKICGLLASEAGCGCKALOLLHESMER 281
Db 243 GWLPDPSEFRRRRAHREKINIFYKAIQKRUSK-----EPAEIIOLLTLDSTYK 292
Oy 282 GERLDMOALKOSSTELFGHETTASATSLITYLGLYPHVLQKVEELKSGLLCKSNQ 341
Db 293 GRPLDDDELTAGMLIGLLLAGOHTSSTTSAWMGFILARKPLQDKCYLBQKT--VCGEDL 349
Oy 342 DNKLDMLELEOLKYIGCVIKETLLNPPVPGGRFVALKTFELNGQIPKGNVYISICT 401
Db 350 P-PLYEOLKDLNLDRCIKETRLRPPIMTMMRAKTPQYVAGTTPGHOVCSPVN 408
Oy 402 HVALEFTNKEEFNDRFSAHPEDASRSFIPFGGRLSCGKEFAKILKIFTVELAR 461
Db 409 QRRKDSWRELDFNDRIQDNPASGEKFAIYPPGAGHRICGENFAIVQIKTMSMLR 468
Oy 462 HCDWOLLNGP-PTMKTSPTVYVDNLPAFE 490
```

Db 469 LYEPDLINGYPSVNTYTHHHPENPVIRX 498

```

RESULT 13
C4D1_DROME STANDARD; PRT; 512 AA.
AC P33269; 0186644; 018653; 09W515; 09W516;
DT 01-FEB-1994 (Rel. 28, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE CYTOCHROME P450 4D1 (EC 1.14.-.-) (CYP1D1).
GN CYP4D1 OR CYP-P450-D1 OR EG:87B1.1 OR CG3656.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=OREGON-R; TISSUE=Embryo;
RC MEDLINE=92297166; PubMed=1605661;
RA Gandhi R., Varak E., Goldberg M.L.;
RT "Molecular analysis of a cytochrome P450 gene of family 4 on the
RL DNA Cell Biol. 11:397-404(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=VARIOUS STRAINS;
RA Phillips K.S., Begun D.J., Aquadro C.F.;
RT "Evidence for non-neutral evolution around the cytochrome P450 gene
RL cluster on the Drosophila melanogaster X chromosome.";
[3]
RN submitted (OCT-1997) to the EMBL/Genbank/DBJ databases.
[4]
RN SEQUENCE FROM N.A.
RP STRAIN=OREGON-R;
RC MEDLINE=20196011; PubMed=10731137;
RA Benoit P.V., Gatt M.K., Ashburner M., Murphy L., Harris D.,
RA Barrett B.G., Ferraz C., Vidal S., Brun C., Demallies J., Borkova D.,
RA Dreano S., Gloux S., Lelaure V., Motlier S., Galbert F., Borkova D.,
RA Minana B., Kafatos F.C., Louis C., Siden-Kiamos I., Bolshakov S.,
RA Papagiannakis G., Spanos L., Cox S., Madueno E., de Pablos B.,
RA Modolell J., Peter A., Schotler P., Werner M., Mourikoti F.,
RA Belmont N., Dove G., Schaffer U., Jackle H., Bucheton A.,
RA Callister D.M., Campbell L.A., Darlamsiou A., Henderson N.S.,
RA McMillan P.J., Sallies C., Tait E.A., Valenti P., Saender R.D.,
RA Glover D.M.;
RT "From sequence to chromosome: the tip of the X chromosome of D.
RL melanogaster.";
[5]
RN Science 287:2220-2222(2000).
[6]
RN SEQUENCE FROM N.A. (LONG AND SHORT ISOFORMS).
RP STRAIN=BERKELEY;
RC MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S.S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang O., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abriil J.F., Agbayani A., An H.-U., Andrews-Pinnock C., Baldwin D.,
RA Balilew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Esler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glaser K.,
RA Glorick A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Hostin D., Havel D., Heiman T.J., Hernandez J.R., Houck C.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegam C.,

```

```

RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Matel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman C.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
CC -!- FUNCTION: INVOLVED IN THE METABOLISM OF INSECT HORMONES AND IN THE
CC BREAKDOWN OF SYNTHETIC INSECTICIDES.
CC -!- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) = ROH +
CC OXIDIZED FLAVOPROTEIN + H(2)O.
CC -!- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM.
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC SHORT FORM; MAY BE PRODUCED BY ALTERNATIVE SPLICING.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED THROUGHOUT DEVELOPMENT, WITH THE
CC HIGHEST LEVELS OCCURRING DURING LATE LARVAL STAGES, THEN FALLING
CC DRASTICALLY DURING PUPARIATION.
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; X67645; CAA47887.1; -
DR EMBL; Z98269; CAB10972.1; -
DR EMBL; AF016992; AAB71155.1; -
DR EMBL; AF016993; AAB71156.1; -
DR EMBL; AF016994; AAB71157.1; -
DR EMBL; AF016995; AAB71158.1; -
DR EMBL; AF016996; AAB71159.1; -
DR EMBL; AF016997; AAB71160.1; -
DR EMBL; AF016998; AAB71161.1; -
DR EMBL; AF016999; AAB71162.1; -
DR EMBL; AF017000; AAB71163.1; -
DR EMBL; AF017001; AAB71164.1; -
DR EMBL; AF017002; AAB71165.1; -
DR EMBL; AF017003; AAB71166.1; -
DR EMBL; AF017004; AAB71167.1; -
DR EMBL; AE003423; AAF45736.1; -
DR EMBL; AE003423; AAF45737.1; -
DR PIR; S25707; S25707; -
DR FlyBase; FBgn0005670; Cyp4d1.
DR InterPro; IPR001128; -
DR InterPro; IPR002402; -
DR Pfam; PF00067; P450; 1.
DR PRINTS; PR00385; P450.
DR PRINTS; PR00464; PF45011.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
KW Oxidoreductase; Monooxygenase; Membrane; Heme; Microsome;
KW Endoplasmic reticulum; Alternative splicing; Polymorphism.
FT BINDING 456 456
FT VASPLIC 1 186
FT -----
FT MVYIGAILASALPVGGLTHLKKRRIIDISYRGPPVLP
FT LVGHGHFIKRPHEWKKIIEPMYRSKQVAKWVGPEL
FT NVLGNKRDVEYVIGTRPNDKACEYKALEPWLKEGLVSR
FT GRKWHKRRKILTPAFHFKIILDQVEYVEKSRDLRWEMD
FT RLKRGESFLYDWININCLMDT -> MMLLSITVLTAITA
FT LEMRRFLRNMTIPGPIPLDGLNAAHIFLGLTFAEACIKIG

```

```

FT FT ELARHDTGTCFLGPGSYWLFNPRVERVIGSSQTLTSS
FT FT QEISFLGRWLNELGLVNSGRKWRKRRIITPAHFRLPEP
FT FT VEIPDRQSLRLVELALRISNGORIRIMGEAHLICALDA
FT FT (IN SHORT ISOFORM).
FT FT VARIANT 170 170 CAM-12, CAM-41, CAM-44, CAM-48 AND
FT FT BERKELEY).
FT FT S -> I (IN STRAINS CAM-8, CAM-44, CAM-48
FT FT AND BERKELEY).
FT FT CONFLICT 68 68 Y -> I (IN REF. 1).
FT FT CONFLICT 316 316 E -> K (IN REF. 1).
FT FT CONFLICT 468 509 AIVANVLRYHVEVDGSEPPVLIATLITRKEPLMEKVR
FT FT E -> PSMFPGSGTTRTLTATSTGTTTRADRTTSATQGLP
FT FT SSRGCG (IN REF. 1).
SQ SEQUENCE 512 AA; 58691 MW; F7B089734231A6A3 CRC64;

Query Match 12.4%; Score 324.5; DB: 1; Length 512;
Best Local Similarity 24.9%; Pred. No. 3.1e-15;
Matches 132; Conservative 83; Mismatches 209; Indels 107; Gaps 20;

OY 5 ALLASALCFVPLLEFLTAIKLMDLYCVRGDRSCALPLPGTMGFPFGETLQMLOR 64
DB 7 ALLASGL--FVGLLLHLKFKRLIDL-----ISYMGPPVPLVY----- 44
OY 65 RKFLQMKRRKRYGYFYKTHLFGPPT-----VRMGADNVRKILLGDRLVSVHMPA 114
DB 45 -----HGHFPIGKPRHEWVKRIPEFMEYSKDDVLVWGLPELNVLMGNK 90
OY 115 SVRTILGS-----GCLSNLHDSHQRKVKIMRAFSRLECYVY 155
DB 91 DVEVVLGTLRFNDKAGEYKALEPWLKEGLISGRKKMK-RKKITTPAFHFILDOFVEY 149
OY 156 ITEEVGS-----SLDQ-WLSCGERGLVYEVYKRLMFRIMRILLGCEPQLAGDSDSEQ- 208
DB 150 F--EKSGRLLNMEQDRKHGESGSLDWINTLCTMDTICETAMGVSNAGSNDSEYV 207
OY 209 QLVNEAF-----BEKTRNLESLPDPFSGLYNG-MKARLILHARIEONT---RAKICGR 259
DB 208 QAVKITSMVLHKRMFNILYRFDLTYMLTFLARAEEKALVNLQFTEKIIIVORREEL--IR 265
OY 260 ASNAGGCKD-----ALQILIEHSWGER-LDMQALKOSTELLFGHETTS 307
DB 266 EGGSSQSSNDADVGAKRMAFLLDILQSTVD--EKPLNLDIREVDTFMFEGHDTTS 323
OY 308 AATSLITLYGLYPHVLOKRYEELSKGLCKSNODNKLDMELIEQLKYIGVYIKETLRIN 367
DB 324 ALMFFPYNIATHEAQKCFEELRS---VVGNDKSPVSYELLNOLHYVDLCVKETLRMY 380
OY 368 PVPYGGFVALKTFEELNGYQIPKGNVYISCDTHDAVEIFTNKEEFNDRFS-APHED 426
DB 381 PSYVLGLRKRLDECEENKGLIPAGTNGISPLYLGRREELFSEPNFKEBEREDVYVTTAK 440
OY 427 ASNFSPFPGGGLRSGVCKEFAILLIKITVELARHCDMLL---NGPPTM 474
DB 441 LNPYAVIIPISAGPRNCTIGQFAMLETKAIVANVLRHYEVDVGDSEPPVL 491

RESULT 14
CP84_ARATH STANDARD: PRT: 520 AA.
AC 042600;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE CYTOCHROME P450 8A41 (FERULATE-5-HYDROXYLASE) (EC 1.14.-.-) (FSH).
GN CYP84A1 OR FAH1 OR AT4G36220 OR F23E13.110.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxId=3702;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=CV, COLUMBIA;
RX MEDLINE=96293440; PubMed=8692910;
RA Meyer K., Cusumano J.C., Somerville C.R., Chapple C.C.S.;
RT "Ferulate-5-hydroxylase from Arabidopsis thaliana defines a new
RL family of cytochrome P450-dependent monooxygenases.";
RL Proc. Natl. Acad. Sci. U.S.A. 93:6869-6874(1996).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, COLUMBIA;
RX MEDLINE=20083488; PubMed=10617198;
RA Meyer K.F.X., Schueller A., Wambutt R., Murphy G., Volkhaert G.,
RA Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terryn N.,
RA Harris B., Ansoorge W., Brandt P., Grivell L., Rieger M.,
RA Weichselgartner M., de Simone V., Obermaier B., Maché R., Mueller M.,
RA Kreis M., Delseny M., Puigdomenech P., Watson M., Schmidheini T.,
RA Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,
RA Vos P., Hohelsel J., Zimmermann W., Wedler H., Ridley P.,
RA Langham S.-A., McCullagh B., Bilham L., Robben J.,
RA Van der Schueren J., Grynoprez B., Chang Y.-J., Vandebussche F.,
RA Braeken M., Weltjens I., Voet M., Bastiaens J., Aert R., Defoor E.,
RA Wiltzenegger T., Bothe G., Rampsperger U., Hilbert H., Braun M.,
RA Holzer E., Brandt A., Peters S., van Staveren M., Dirkse W.,
RA Moolman P., Klein Lankhorst R., Rose M., Haut J., Koetter P.,
RA Berneser S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,
RA De Keyser A., Buyschaert C., Gielen J., Villarroel R., De Clercq R.,
RA Van Koutgou M., Rogers J., Cronin A., Quail M., Bray-Allen S.,
RA Clark L., Doggett J., Hall S., Kay M., Lennard N., McIay K., Mayes R.,
RA Petteit A., Rajandream M.-A., Lyne M., Benes V., Reckmann S.,
RA Botkova D., Blocker H., Scharte M., Grimm M., Loehner T.-H.,
RA Godel S., de Haan M., Maarse A., Schaefer M., Mueller-Auer S.,
RA Babel C., Fuchs M., Fartman B., Grandérath K., Dauner D., Herzl A.,
RA Neumann S., Argirion A., Vitale D., Liquri R., Pivrandi E.,
RA Massenet O., Quigley F., Clabaud G., Wendlein A., Felder R.,
RA Schnabl S., Hiller R., Schmidt W., Lecharry A., Aubourg S.,
RA Chedfor F., Cooke R., Berger C., Monfort A., Casacuberta E.,
RA Gibbons T., Weber N., Vandenbol M., Baryes M., Terol J., Torres A.,
RA Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,
RA Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Bielek C.,
RA Frishman D., Haase D., Lemcke K., Kewes H.-W., Stocker S.,
RA Zaccaria P., Bavan M., Wilson R.K., de la Bastide M., Hebermann K.,
RA Parnell L., Dedha N., Gnoj L., Schutz K., Huang E., Spiegel L.,
RA Selkon M., Murray J., Sheet P., Cordes M., Abut-Friedel J.,
RA Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,
RA Latelle P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
RA Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
RA Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,
RA Nelson J., Spieth J., Ryan E., Andrews S., Geisel C., Layman D.,
RA Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshi C.,
RA Antonoli B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,
RA Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shah R.,
RA Swaby I.K., O'Shaugnessy A., Rodriguez M., Hoffman J., Till S.,
RA Grant S., Shohdy N., Hasegawa A., Hamed A., Lodi M., Johnson A.,
RA Chen E., Marra M., Martensen R., McCombie W.R.;
RT "Sequence and analysis of chromosome 4 of the plant Arabidopsis
RL thaliana.";
RL Nature 402:769-777(1999).
[3]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, LANDSBERG ERCTA;
RX MEDLINE=99097044; PubMed=9880351;
RA Ruegger M., Meyer K., Cusumano J.C., Chapple C.;
RT "The regulation of ferulate-5-hydroxylase expression in Arabidopsis in
RL the context of sinapate ester biosynthesis.";
RL Plant Physiol. 119:101-110(1999).
-1- PATHWAY: GENERAL PHENYLPROPANOID PATHWAY.
-1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```


Wed Nov 7 09:29:25 2001

us-09-668-482-4.rsp

Page 13

Search completed: November 6, 2001, 13:39:40
Job time: 243 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2001, 13:38:25 ; Search time 115.92 Seconds

(without alignments)
567.250 Million cell updates/sec

Title: US-09-668-482-4

Perfect score: 2611
1 MGLPALLASALCTFVLPPLL.....PTVFPVDNLPAHFTHFGEI 497

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :
1: SPREMBL.16:*
2: sp_archaea:*
3: sp_bacteria:*
4: sp_fungi:*
5: sp_human:*
6: sp_invertebrate:*
7: sp_mhc:*
8: sp_mammal:*
9: sp_organelle:*
10: sp_phage:*
11: sp_plant:*
12: sp_rodent:*
13: sp_unclassified:*
14: sp_vertebrate:*
15: sp_virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2434	93.2	497	11 Q9RIF4	Q9RIF4 mus musculu
2	2030.5	77.8	492	13 Q9PUB4	Q9PUB4 gallus gall
3	1739.5	66.6	492	13 Q93323	Q93323 xenopus lae
4	1041.5	39.9	512	4 Q9NR63	Q9NR63 homo sapien
5	1029	39.4	525	4 Q9NP41	Q9NP41 homo sapien
6	663.5	25.4	444	2 Q59990	Q59990 synechocyst
7	565	21.6	477	10 Q9LWY7	Q9LWY7 arabisidops
8	535	20.5	107	13 Q9PUG2	Q9PUG2 gallus gall
9	519.5	19.9	485	10 Q9SJH2	Q9SJH2 arabisidops
10	517.5	19.8	463	10 Q9FV76	Q9FV76 arabisidops
11	512	19.6	457	10 Q65624	Q65624 arabisidops
12	492.5	18.9	482	10 Q81077	Q81077 arabisidops
13	450.5	17.3	465	10 Q9FMA5	Q9FMA5 arabisidops
14	439	16.8	513	10 Q64989	Q64989 arabisidops
15	439	16.8	513	10 Q9SC09	Q9SC09 arabisidops
16	436	16.7	443	10 Q9LJK2	Q9LJK2 arabisidops
17	434	16.6	489	10 Q9ZV72	Q9ZV72 arabisidops
18	426	16.3	474	10 Q9LKH7	Q9LKH7 phaseolus a
19	426	16.3	478	10 Q9LNT3	Q9LNT3 arabisidops

20	417.5	16.0	496	10 Q9FOX4	Q9FOX4 cucurbita m
21	411.5	15.8	465	10 Q9LH81	Q9LH81 arabisidops
22	409	15.7	504	10 Q9SNG3	Q9SNG3 oryza sativ
23	406	15.5	457	10 Q9M066	Q9M066 arabisidops
24	406	15.5	524	10 Q23242	Q23242 arabisidops
25	391.5	15.0	464	10 Q9LIC5	Q9LIC5 arabisidops
26	377	14.4	475	2 Q9X7G9	Q9X7G9 myxococcus
27	348.5	13.3	511	10 Q9SWR1	Q9SWR1 liquidambar
28	347	13.3	486	11 Q9J1Y3	Q9J1Y3 mus musculu
29	347	13.3	503	11 Q9J1P8	Q9J1P8 mus musculu
30	344	13.2	318	10 Q9LVP3	Q9LVP3 arabisidops
31	338	12.9	521	10 Q9XFM2	Q9XFM2 lycopersico
32	335.5	12.8	513	10 Q9FVB8	Q9FVB8 brassica na
33	334	12.8	518	10 Q9F1I8	Q9F1I8 arabisidops
34	333.5	12.8	512	10 Q9FX29	Q9FX29 arabisidops
35	333.5	12.8	520	10 Q9FVC0	Q9FVC0 brassica na
36	332	12.7	520	10 Q9FVB9	Q9FVB9 brassica na
37	328.5	12.6	474	10 Q9SHY7	Q9SHY7 arabisidops
38	327	12.5	498	5 O16806	O16806 drosophila
39	326	12.5	464	10 Q04949	Q04949 arabisidops
40	318.5	12.2	520	10 Q48786	Q48786 arabisidops
41	318	12.2	503	5 Q21424	Q21424 caenorhabdi
42	317.5	12.2	735	10 Q9LGI7	Q9LGI7 arabisidops
43	313.5	12.0	488	10 Q9SAA9	Q9SAA9 arabisidops
44	313.5	12.0	513	10 Q9SBP8	Q9SBP8 populus tri
45	312.5	12.0	491	5 Q9NGK3	Q9NGK3 tribolium c

ALIGNMENTS

RESULT 1

Q9RIF4 PRELIMINARY: PRT; 497 AA.

AC Q9RIF4;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE CYTOCHROME P450 RETINOIC ACID METABOLIZING ENZYME P450RA.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=ROOTH;
RX PubMed=11063033;
RA Paine C.T., Paine M.L., Snead M.L.;
RT "Identification of tufelin- and amelogenin-interacting proteins using
RT the yeast two-hybrid system.";
RL Connect. Tissue Res. 38:257-267(1998).
CC -1- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) = ROH +
CC OXIDIZED FLAVOPROTEIN + H(2)O (By SIMILARITY).
CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM (By
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC EMBL: AF115769; AAD17217.1; -.
DR InterPro: IPR001128; -.
DR Pfam: PF00067; P450; 1.
DR PRINTS: PR00385; P450.
DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
KW Electron transport; Endoplasmic reticulum; Heme; Membrane; Microsome;
KW Monooxygenase; Oxidoreductase.
SQ SEQUENCE 497 AA; 56191 MW; 67CB167A997842C1 CRC64;

Query Match 93.2%; Score 2434; DB 11; Length 497;

Best Local Similarity 93.2%; Pred. No. 1.4e-190;
Matches 463; Conservative 15; Mismatches 19; Indels 0; Gaps 0;

QY 1 MGLPALLASALCTFVLPPLLFLAIKIMLDYCVSGRDSALPLPPTMGPPFGETLQM 60
DB 1 MGLPALLATATCTFVLPPLLFLAALKIMLDYCVSSRDSALPLPPTMGPPFGETLQM 60

OY	61	VLORRKFLOMRKRGKGFYKTHLHGRPRVRMGADNRRILLGGDRDLVSVHWPASVITL	120
OY	61	VLORRKFLOMRKRGKGFYKTHLHGRPRVRMGADNRRILLGGDRDLVSVHWPASVITL	120
Db	61	VLORRKFLOMRKRGKGFYKTHLHGRPRVRMGADNRRILLGGDRDLVSVHWPASVITL	120
OY	121	GGSCSLNLDHSSHQOKKVKVIMRAFSSREALTECVYVITEEVSSSLEQWLSGGERGLTYPE	180
Db	121	GAGCSLNLHSSSHQOKKVKVIMRAFSSREALTECVYVITEEVSSSLEQWLSGGERGLTYPE	180
OY	181	VKRLMFRIAMTLLIGCEPQLAGDDSEQOLYEAFEEBTRNLFSLPIDVPSSGLRGKMKAR	240
Db	181	VKRLMFRIAMTLLIGCEPQLAGDDSEQOLYEAFEEBTRNLFSLPIDVPSSGLRGKMKAR	240
OY	241	NLIHARIQNTIRAKICGRVASEAOGCKDALOLIEHSWSEGERELDMQALKOSTELLFG	300
Db	241	NLIHARIQNTIRAKIRLQATPEPPGCGCKDALQILLIEHSWSEGERELDMQALKOSTELLFG	300
OY	301	GHETTASATSLTYLGLYPRVLOKVRBELKSKGLCKSNODKMLDEILLOKTYIGCVI	360
Db	301	GHETTASATSLTYLGLYPRVLOKVRBELKSKGLCKSNODKMLDEILLOKTYIGCVI	360
OY	361	KETTLRLNPVVGGRFVALLKPELNGYOLIPGMWNTYSICDTHVAELFTTKKEFNDRFS	420
Db	361	KETTLRLNPVVGGRFVALLKPELNGYOLIPGMWNTYSICDTHVAELFTTKKEFNDRFL	420
OY	421	APHPEDASRFSTPEPGGGLRSCVCKEERAKILTKIFYTELARHCDWQILNPPMKTSPTV	480
Db	421	VPHPEDASRFSTPEPGGGLRSCVCKEERAKILTKIFYTELARHCDWQILNPPMKTSPTV	480
OY	481	YPVDNLPARFTHGEEL 497	
Db	481	YPVDNLPARFTHGEEDI 497	
RESULT 2			
O9PUB4			
ID	O9PUB4	PRELIMINARY;	PRT; 492 AA.
AC	O9PUB4		
DT	01-MAY-2000 (TREMBLrel. 13, Created)		
DT	01-MAY-2000 (TREMBLrel. 13, last sequence update)		
DT	01-MAR-2001 (TREMBLrel. 16, last annotation update)		
DE	CYTOCROME P450 26 (EC 1.14.-.-) (RETINOIC ACID-METABOLIZING		
DE	CYTOCROME) (P450RA1) (RETINOIC ACID 4-HYDROXYLASE).		
GN	CYP26A1.		
OS	Gallus gallus (Chicken).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;		
OC	Gallus.		
NC	NCBI_TaxID=9031;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Swinebel E.C., Thaller C., Sockanathan S., Petkovich M., Jessell T.M.		
RA	Etchele G.,		
RT	"Complementary Domains of Retinoic Acid Production and Degradation in		
RT	the Early Chick Embryo."		
RL	Dev. Biol. 0:0-0(2000).		
CC	-1- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) = ROH +		
CC	OXIDIZED FLAVOPROTEIN + H(2)O (BY SIMILARITY).		
CC	-1- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM (BY		
CC	SIMILARITY).		
CC	-1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.		
DR	EMBL: AF199462; AAF09250.1; -.		
DR	Interpro: IPR001128; -.		
DR	Pfam: PF00067; P450; 1.		
DR	PRINTS: PR00385; P450.		
DR	PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.		
KW	Electron transport; Endoplasmic reticulum; Heme; Membrane; Microsome;		
KW	Monooxygenase; Oxidoreductase.		
QO	SeqUnce 492 AA; 5526 MW; 7f28b72ef5c232fb CRC64;		

Query Match	77.8%;	Score 2030.5;	DB 13;	Length 492;
Best Local Similarity	78.9%;	Pred. NO. 1.3e-157;		

Matches	392;	Conservative	43;	Mismatches	57;	Indels	5;	Gaps	3;
---------	------	--------------	-----	------------	-----	--------	----	------	----

QY	1	MCPIPLLSALCTEFLVPLILFLPLAATKIIMDIYCVSRRBSGALPLPCGMGPPFGEFLQM	60
Db	1	MFPSALVNSALCTEFLPLILFLPLAARIMDLVCASGRDSCPLPLPCGMGPPFGEFLQM	60
QY	61	VLQRRKFLQMKRRKKGFLYKTHLEFGPRVRYMGADNVARRILGLDRLVSYVMPASVPTIL	120
Db	61	VLQRRKFLQMKRRKKGFLYKTHLEFGPRVRYMGADNVARRILGLDRLVSYVMPASVPTIL	120
QY	121	GGGCLSNLHDSHKKRKYVIMAFSREALECYVPVITEEYVSSLEQMLSCGERGLVYPE	180
Db	121	GGGCLSNLHDSHKKRKYVIMAFSREALECYVPVITEEYVSSLEQMLSCGERGLVYPE	180
QY	181	YRLMFRIMARLILGCEPQLAGDGSBQOLYEAEBEMTRNLSPIPIVPPSGLYRGKKAR	240
Db	180	YRLMFRIMARLILGCEPQLAGDGSBQOLYEAEBEMTRNLSPIPIVPPSGLYRGKKAR	238
QY	241	NLIHARIPONIAKICGIRASBAGGCCDAÖLLIHSWMEGERLDMQALKOSTELLFG	300
Db	239	NLIHAKIEENIRAK---NAKREPEGYKDALÖLLMERTÖGNGEÖLNMÖLKESATELLFG	295
QY	301	GHETTASATSLITLYGLYPHYLOKVRBELSKYLCKSNODNKLMDEILÖÖLYICCVI	360
Db	296	GHETTASATSLIARLGLHNDVLQVKREQLQKGLSGPNÖKÖLNNEPLÖÖLYKTCVIL	355
QY	361	KETLILNPPYVGGFVALKTELNGYÖLPKGMNVIYSICDTHVAELFTKKEFPNDRFS	420
Db	356	KETLILSPYPVGEGFIAKTELNGYÖLPKGMNVIYSICDTHVADELFTKDEFNDRFM	415
QY	421	APHPDARFSFPIPGGGLRSCVCKEPAKILKLFYVILARHCÖMÖLLNPPIMTKSPV	480
Db	416	SSPSPDSSRFSPPIPGGGLRSCVCKEPAKILKLFYVILARSCÖMÖLLNPPIMTKGPV	475
QY	481	YPVDMLPARFTFHGEI 497	
Db	476	YPVDMLPARFTFHGEI 492	
RESULT	3		
093323			
AC	093323	PRELIMINARY;	PRT; 492 AA.
DT	01-NOV-1998 (TReMBLrel. 08, Created)		
DT	01-NOV-1998 (TReMBLrel. 08, Last sequence update)		
DT	01-MAR-2001 (TReMBLrel. 16, Last annotation update)		
DE	CYTÖCHROME P450 26 (EC 1.14.-.-) (RETINOIC ACID-METABOLIZING		
DE	CYTÖCHROME (P450RA1) (RETINOIC ACID 4-HYDROXYLASE).		
OS	Xenopus laevis (African clawed frog).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipidae;		
OC	Xenopodinae; Xenopus.		
OX	NCBI_TaxID=8355;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Chen J.L., Grunz H., Panitz F., Pieler T., Hollmann T.;		
RL	Submitted (APR-1998) to the EMBL/GenBank/DBD databases.		
CC	-1- SIMILARITY: BELONGS TO THE CYTÖCHROME P450 FAMILY.		
DR	EMBL; AF057566; AAC25158.1; -		
DR	InterPro; IPR001128; -		
DR	Pfam; PF00067; P450; 1.		
DR	PRINTS; PR00385; P450.		
DR	PROSITE; PS00086; CYTÖCHROME_P450; UNKNOWN_1.		
FM	Heme; Monooxygenase; Oxidoreductase.		
QW	SEQUENCE 492 AA; 55459 MW; DID4BB7651BF2D3E CRC64;		

Query Match	66.6%	Score 1739.5;	DB 13;	Length 492;
Best Local Similarity	67.1%;	Pred. No. 7.5e-134;		

Qy 1 MGLPALIASALCTEVPDLLFLTAIKLMDLVCVSGRDRCALPLPGTGMGFPFGETLQW 60

Db 1 MDYTLTSLALCTIALPULLLLLFPAKIMEVYCLRRKDAACANPLPFGTGLPFGERTLQM 60
 QY 61 VLQRRRLQMKRRRYGFTYTHLFGPPTVVMGADNVRILLGDDRLVSVHPASVTTIL 120
 Db 61 VLQRRRLQMKRRRYGFTYTHLFGPPTVVMGADNVRILLGDDRLVSVHPASVTTIL 120
 QY 121 GSGGLSNLHSHSHQKRRKVMBAFSEALCYVPVITEEYSSLSLEQLSGCERGLLVYPE 180
 Db 121 GAGGLSNLHSHSHQKRRKVMBAFSEALCYVPVITEEYSSLSLEQLSGCERGLLVYPE 179
 QY 181 VKRLMFRIAMRILLGCEPQIAGDSEQLVEAFEEEMTRNLSLPIDVPSGLYRGKAR 240
 Db 180 IKRMFRIRAMLLGCGPQRM-DREGEETILLEAFEEEMSRNLSLPIDVPSGLYRGKAR 238
 QY 241 NLHARIEQNRIRAKICGLRASEAGGCKDALQILLIEHSEMERGERLDMQALKOSTELLFG 300
 Db 239 NLHARIEQNRIRAKICGLRASEAGGCKDALQILLIEHSEMERGERLDMQALKOSTELLFG 294
 QY 301 GHETTASATSLITYLGLYPHVLOKREELKSKGLL-CKSNQDKLDMETLEQLKYIGV 359
 Db 295 GHETTASATSLITYLGLYPHVLOKREELKSKGLL-CKSNQDKLDMETLEQLKYIGV 354
 QY 360 IKETRLNPPVPGGFRVALKTFEELNGYQIPKGMNVIYSICDTHVAEFTNKEEFNDRF 419
 Db 355 IKETRLNPPVPGGFRVALKTFEELNGYQIPKGMNVIYSICDTHVAEFTNKEEFNDRF 414
 QY 420 SAPHEDEASRESFIPFGGGLRSCVGEKFAKILKIFVEELARHCDWOLLNCPPTMKTSPT 479
 Db 415 LTPPLRDSRSGFIPFGGGLRSCVGEKFAKILKIFVEELARHCDWOLLNCPPTMKTSPT 474
 QY 480 VYPVNDLPARFTHFGEI 497
 Db 475 ICPVNDLPARFTHFGEI 492

RESULT 4
 ID Q9NR63 PRELIMINARY; PRT; 512 AA.
 AC Q9NR63; 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DE 01-MAR-2001 (Tremblrel. 16, Last annotation update)
 DE CYTOCHROME P450 RETINOID METABOLIZING PROTEIN P450RAI-2.
 GN CYP26B1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP MEDLINE FROM N.A.
 RX MEDLINE-20300913; PubMed-10823918;
 RA White J.A., Ramshaw H., Taimi M., Stangle W., Zhang A., Everingham S.,
 RA Creighton S., Tam S.-P., Jones G., Petkovich M.,
 RA "Identification of the human cytochrome P450, P450RAI-2, which is
 RA predominantly expressed in the adult cerebellum and is responsible for
 RA all-trans-retinoic acid metabolism.";
 RT Proc. Natl. Acad. Sci. U.S.A. 97:6403-6408(2000).
 RL -1- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) = ROH +
 CC OXIDIZED FLAVOPROTEIN + H(2)O (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM (BY
 CC SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 DR EMBL; AF252297; AAF76003.1; -.
 DR InterPro; IPR001128; -.
 DR Pfam; PF00067; P450; 2.
 DR PRINTS; PR00385; P450.
 DR PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_1.
 DR Election transport; Endoplasmic reticulum; Heme; Membrane; Microsome;
 KW Monooxygenase; Oxidoreductase.
 SQ SEQUENCE 512 AA; 57512 MW; A06DD1944E6726F CMC64;

Query Match 39.9%; Score 1041.5; DB 4; Length 512;

Best Local Similarity 44.2%; Pred. No. 8.4e-77;
 Matches 219; Conservative 88; Mismatches 165; Indels 23; Gaps 9;
 QY 7 LASALCTEVLPL-----LFLAIAIKMDLYCVSGRDRSCALPLPFGMGPPFGERTLQMWL 62
 Db 8 LVSALTLAACLVSVTILLAVSQQLQRMATRDSCSLPIPKSGMGPFLIGEGHMLL 67
 QY 63 QRRFLQMKRRKYGFTYKTHLFGPPTVVMGADNVRILLGDDRLVSVHPASVTTILGS 122
 Db 68 QSGGFQSSRRREKYGFTYKTHLFGPPTVVMGADNVRILLGDDRLVSVHPASVTTILGCP 127
 QY 123 GGLSNLHSHSHQKRRKVMBAFSEALCYVPVITEEYSSLSLEQLSGCERGLLVYPEVK 182
 Db 128 NIVNSLIGDHRKRRKRVESKIFSEALLESYLPKIQVLDTLRAMSHDE-AINYQEQAO 186
 QY 183 RLMFRIAMRILLGCE-PQIAGDSEQLVEAFEEEMTRNLSLPIDVPSGLYRGKARN 241
 Db 187 KLFRRMIRVLLGFSIDE-----EDGLHFEVYQFVVDVFSLPVDLPSPGYRRGIQAO 241
 QY 242 LHHARIEQNRIRAKICGLRASEAGGCKDALQILLIEHSEMERGERLDMQALKOSTELLFG 301
 Db 242 LHKGLEKAIKREKL-----OCTGKDYLDLIDLLESSEKHEKMTQELKQGLLEIFPA 297
 QY 302 HETTASATSLITYLGLYPHVLOKREELKSKGLL-CKSNQDKLDMETLEQLKYIG 357
 Db 298 YATTASATSLIMQLKHPVLEKLDLRAHGLIHSGCCPC--EGTLRLDITLSGLRYID 355
 QY 358 CVIKETRLNPPVPGGFRVALKTFEELNGYQIPKGMNVIYSICDTHVAEFTNKEEFND 417
 Db 356 CVIKETRLNPPVPGGFRVALKTFEELNGYQIPKGMNVIYSICDTHVAEFTNKEEFND 415
 QY 418 FESAPHEDE-SRFSFIPFGGGLRSCVGEKFAKILKIFVEELARHCDWOLLNCP-PTMK 475
 Db 416 FESQARSEDKGRHHTVFPFGGVRTCLGKHLAKLFLKVLAVELASISREPLATRTPTRI 475
 QY 476 TSPVYVNDLPARF 490
 Db 476 LVPYLPHPVDGLSVKF 490

RESULT 5
 ID Q9NP41 PRELIMINARY; PRT; 525 AA.
 AC Q9NP41; 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DE 01-MAR-2001 (Tremblrel. 16, Last annotation update)
 DE WUGSC:H_NH0493L16.1 PROTEIN.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP MEDLINE FROM N.A.
 RX MEDLINE-99063792; PubMed-9847074;
 RA Sulston J.E., Waterston R.,
 RA "Toward a complete human genome sequence.";
 RT Genome Res. 8:1097-1108(1998).
 RL [2]
 RP SEQUENCE FROM N.A.
 RA Bourne S., Bauer C., Page K., Jones T.; RP11-493L16.";
 RT "The sequence of Homo sapiens BAC clone genome RP11-493L16.";
 RT Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Waterston R.;
 RA Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) = ROH +
 CC OXIDIZED FLAVOPROTEIN + H(2)O (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM (BY
 CC SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 DR EMBL; AC007002; AAF65576.1; -.

Query Match	Best local similarity	Score	DB	Length	Matches	Conservative	Mismatches	Indels	Gaps
21.6%	31.2%	565	10	477	154	82	209	48	15
<p>NCBI_TaxId=3702:</p> <p>RP SEQUENCE FROM N.A.</p> <p>RC STRAIN=COLUMBIA.</p> <p>RX MEDLINE=20181125; PubMed=10718197;</p> <p>RA Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Kotani H., Tabata S.;</p> <p>RT "Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence features of the regions of 3,076,755 bp covered by sixty P1 and TAC clones."</p> <p>RU DNA Res. 7:31-63(2000).</p> <p>CC -1- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) = ROH + OXIDIZED FLAVOPROTEIN + H(2)O (5% SIMILARITY).</p> <p>CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM (BY SIMILARITY).</p> <p>CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY. -</p> <p>DR EMBL: AB018112; BAA96885.1; -.</p> <p>DR InterPro: IPR001128; -.</p> <p>DR Pfam: PF00067; P450; 1.</p> <p>DR PROSITE: PRO0385; P450.</p> <p>DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.</p> <p>KW Electron transport; Endoplasmic reticulum; Heme; Membrane; Microsome; Monooxygenase; Oxidoreductase.</p> <p>SO SEQUENCE 477 AA; 54850 MW; 304BAB2C4970E405 CnC64.</p>									
Query Match	Best local similarity	Score	DB	Length	Matches	Conservative	Mismatches	Indels	Gaps
21.6%	31.2%	565	10	477	154	82	209	48	15
<p>18 LLLFLAATKMDLYCYSGRDSALP-LPPTGMPFFGETQIAVLQRR-----KFLQMK 71</p> <p>7 ILLFSLILLKRLNLSHFSPYPLPGQTGLPLIGSESFSLASAGQHPKEFTDR 66</p> <p>72 RRYK----GFIKTHLFRTPTVGVGADNVRRILLGDRLVSVHNPASVRYTLGSGCLSN 127</p> <p>67 VRRESSSSSCVFPHLPSPLAVYVGAAGNKFLFTNEKNLVSWMPDSYVKLFPSSMOTS 126</p> <p>128 LHDSSHKQPKVIMAFSR-EALCEYVPVITTEVGSSLE-OWLSGSGRLVYPEYKRLM 185</p> <p>127 SKEARKLR--MLLSQFMKPEALRRYVGMEDIAORHETEW--ANODQVIVFPLTKFT 182</p> <p>186 FRIARILLGGEPOLAGGDSQCLVEAFEEETRLNLSPLIDVPVSGLRGKKARNLDA 245</p> <p>183 FSIARSLTSM-----DPARYKOLEOPNTYAVGIFSLPDLPGTRFKRAIKASNLK 237</p> <p>246 RIEQIRAKIGGLRASEAGOGCKDALDILLIHS-----WEGERLDLMALKOSTELLE 299</p> <p>238 EYSAIVRQRKEELKAGKA-----LEHDDILSHMLNIGETKD-EDLADKITGLI 286</p> <p>300 GGHETTASAASLTLYLGLYPLVLOKVRBELSKGLCKSSNDKNKIDMLTDQLKIGCY 359</p> <p>287 GGHDAVASIVCFVVVNYLAEFPHYORVLOE--QKELIKKKKEKKEGLRWEDIEKMRYSMV 344</p> <p>360 IKETRLNPPVGGGRVALKTFELNGYOIPKGMVNYISCDTHDAVETPTNKEEFPDRF 419</p> <p>345 ACEVWKRIYPPLEGTRREALDHSFGFTYIPKCKMKLWSATATAHMNDYPRPERFEPPNF 404</p> <p>420 --SADPHEDASRFSIPROGGGLRSCVGEKFAKILKIFTELVAHRCOWD-LINGPPTMKT 476</p> <p>405 EESGSKP-----YTFVPPFGGPRMCPKEKAYARLELILIFMNLVNRFKMKVPPENENKIV 459</p> <p>477 SPTVYVVDLPLAR 489</p> <p>460 DPLPIPDKGLPIR 472</p>									
RESULT 8	ID 09PUG2	PRELIMINARY:	PRT: 107 AA.						
AC 09PUG2:									
DT 01-MAY-2000 (Tremblrel. 13, Created)									
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)									

PT	01-MAR-2001 (TREMblrel. 16, Last annotation update)
DE	CYP26 (FRAGMENT).
GN	CYP26.
OS	Gallus gallus (Chicken).
OS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi;
OC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC	Gallus.
OX	NCBI_TaxId=9031;
RN	[1]
RP	SEQUENCE FROM N.A.
RA	Martinez-Ceballos E., Burdjal C.A.;
RL	Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
CC	-1 CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) = ROH +
CC	OXIDIZED FLAVOPROTEIN + H(2O) (BY SIMILARITY).
CC	-1 SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM (BY
CC	SIMILARITY).
CC	-1 SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR	EMBL: AF198526; AAD56546.1; -.
DR	HSSP: P14779; 1B07.
DR	InterPro: IPR001128; -.
DR	Pfam: PF00067; P450; 1.
DR	PRINTS: PR00385; P450.
DR	PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
KW	Electron transport; Endoplasmic reticulum; Heme; Membrane; Microsome;
KW	Monooxygenase; Oxidoreductase.
FT	NON TER
FT	1 107
FT	1 107
SQ	SEQUENCE 107 AA; 12009 MW; CA68BFB0598BC474 CRC64;

	Query Match	20.5%	Score 535	DB 13	Length 107	
	Best Local Similarity	89.7%	Pred. No. 2,1e-36			
	Matches 96	Conservative 7	Mismatches 4	Indels 0	Gaps 0	
OY	373	GERVALKTFEELNGYOIPKGMWNYISICDTHVAETFTNKEEFPDRFSAPHEDEASRSE	432			
DB	1	GERVALKTFEELNGYOIPKGMWNYISICDTHVAETFTNKEEFPDRFSAPHEDEASRSE	60			
OY	433	IPFGGILRSCVCKEFAKILIKFTFELARHCOMOLNPPPTKTSPT	479			
DB	61	IPFGGILRSCVCKEFAKILIKFTFELARHCOMOLNPPPTKTSPT	107			
RESULT	9					
ID	09SJH2	PRELIMINARY:	PRT:	485 AA.		
AC	09SJH2					
DT	01-MAY-2000	(TREMBLrel. 13, Created)				
DT	01-MAY-2000	(TREMBLrel. 13, Last sequence update)				
DT	01-MAR-2001	(TREMBLrel. 16, Last annotation update)				
DE	PUTATIVE CYTOCHROME P450.					
CN	AT2G42850.					
OS	Arabidopsis thaliana (Mouse-ear cress).					
OC	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;					
OC	Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;					
OC	Brassicales; Brassicaceae; Arabidopsis.					
OX	NCBI_TaxID=3702;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RC	STRAIN=CV, COLUMBIA:					
RX	MEDLINE=20083487; PubMed=10617197;					
RA	Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,					
RA	Fuji C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,					
RA	Buell C.R., Ketchum K.A., Lee J.J., Ronald C.M., Koo H., Goffat K.S.,					
RA	Crozin L.A., Shen M., Vanaken S.E., Umayam L., Tallon L.J., Gill J.E.,					
RA	Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,					
RA	Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,					
RA	Salisbury S.L., Fraser C.M., Venter J.C.;					
RT	"Sequence and analysis of chromosome II of Arabidopsis thaliana.";					
RL	Nature 402:761-768(1999).					
CC	-1- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) = ROH +					
CC	OXIDIZED FLAVOPROTEIN + H(2)O (BY SIMILARITY).					
CC	-1- SUBCELLULAR LOCATION: MEMBRAN-BOUND. ENDOPLASMIC RETICULUM (BY					

CC SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 DR EMBL: AC006931; AAD21724.1; -.
 DR InterPro: IPR001128;
 DR Pfam: PF00067; P450; 2.
 DR PRINTS: PR00385; P450.
 DR Electrophoresis: CYTOCHROME_P450; UNKNOWN_1.
 KW Electron transport; Endoplasmic reticulum; Heme; Membrane; Microsome;
 Monooxygenase; Oxidoreductase.
 SQ SEQUENCE 485 AA; 55405 MW; 08B0B16474620E82 CRC64;

Query Match 19.9%; Score 519.5; DB 10; Length 485;
 Best Local Similarity 28.3%; Pred. No. 3,4e-34;
 Matches 132; Conservative 86; Mismatches 201; Indels 47; Gaps 10;

QY 44 LPPGIMGPFPGGELQWVLOKRR-----FLQKKRRKYGITYTHLFGRPVRYMGADNV 97
 DB 45 LPPGEGGLPPIGIMTDFYKQKSNRVEDEVNPRILKHGNIFFKTRIMGSPFIYVNGAEAN 104
 QY 98 RRIILGDDRLVSVHMPASVYTIISGCLSNLHDSHKORKKVMRARSREALCEYVIT 157
 DB 105 RLISMEFSLVSWSSVQMGKMCIMAKGCKHRVLRGIYANSLTIGLESILPKLC 164
 QY 158 EYVGSSE-QLMSCGERGLIYPEVKRKLAFRIAMRILGCEPOLAGDSEQDLVEAFEE 216
 DB 165 DTVKPFHETEMR--GKEEISLYRSKAVLFTVVVECLYGIKEIG-----MLEVEER 214
 QY 217 MFRNLPSLPDVPFSGLYRGMKARNLIHRIEONIRAKICGLRASEAGCKDALQILLIE 276
 DB 215 VLEGVALPPEFPCSFARKKARLEIFLVGRKREK---RREMEKEAEKPNITLF- 269
 QY 277 HSWMERGERLDMQAKOSPE-----DLFGHETTASAASTSLTYLLGYPHVLOKVR 327
 DB 270 -----SRIVEELIKGYTBEVVDNMVLLVFAHDTTSYAMSTMTEFMAQHP-----TCR 319
 QY 328 EELKSGKGLCKSN--QDNKIDMELIQKIGYIKETLFLANPVGGEFVALKTEPLNG 365
 DB 320 DTLQHAQIKAKKGEGLYTVEDVKRMKSMQVYRITMLSPPIGCSFRKAADIDYCG 379
 QY 386 YQIPKGMNIVYSICDTHDVAEIFTNKEEFNDRFSAPHPEDASRFSPIPGGGLRSCVKG 445
 DB 380 YTIIPKGMKILMTYGHYHNEIFQDPMSPDPTFDKP-----IQAYITLPPGGGPRCLAGH 435
 QY 446 ERAKILLKIFVVELARHCWQLLNGPPTKTSPTVYVVDNLPAKPT 491
 DB 436 QIAKISILVEMHFVVTGFMSTLVYPDETISMDDLPPSLCPIKIS 481

RESULT 10
 Q9FH76 PRELIMINARY; PRT; 463 AA.
 AC Q9FH76;
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE CYTOCHROME P450.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 OC Brassicales; Brassicaceae; Arabidopsis.
 NCBI_TaxID=3702;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=COLUMBIA;
 RX MEDLINE=20181125; PubMed=10718197;
 RA Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Kotani H.,
 RA Tabata S.,
 RT Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence
 RT features of the regions of 3,076,755 bp covered by sixty pl and TAC
 RT clones.";
 RL DNA Res. 7:31-63(2000).
 DR EMBL: AB020744; BAB10255.1; -.

SQ SEQUENCE 463 AA; 52366 MW; CCD17293F553F812 CRC64;

Query Match 19.8%; Score 517.5; DB 10; Length 463;
 Best Local Similarity 30.0%; Pred. No. 4,7e-34;
 Matches 145; Conservative 82; Mismatches 220; Indels 37; Gaps 11;

QY 6 LWSALCTFVLPIILFLFAIKIMDLVCSGRDSCALPDPGTMGPFPGGELQWVLO-R 64
 DB 8 LTVSAALF-LCLRFYAGVR-----RSSSTKLPLPFGTMGYVGETFOLYSODP 57
 QY 65 RKFLOMKRRKRYGITYKTHLFGRPVRYMGADNVRIILGDDRLVSVHMPASVYTIISGSC 124
 DB 58 NVFEPAAKQRRYGSYFKTHVLGCPVMISSPEAKFVLVTSHLFKPIFPASKERMLCKQA 117
 QY 125 LSNLHDSHKORKKVMRARSREALCEYVITYEEVGSSLEQMLSCGERGLIYPEVKRL 184
 DB 118 IFPHQGDYHSKRLKRLVLRAMPDAIRMVPHIISIAOESLSNSW---DGTOLNTYQEMKTY 174
 QY 185 MFRIAMRILGCEPOLAGDSEQDLVEAFEEEMTRNLSPIIDVPFSGLYRGMKARNLIH 244
 DB 175 TFVALISILGKD-----EYTRDELKRCYIILEKGNMSPINLPGLFKAKARKREL- 228
 QY 245 ARLEQNIARATIGLRASEAGCKDALQILLIEHSWMERGERLDMQAKOSSTELLFGGHET 304
 DB 229 AQILANI-----LSKRRQNSHTDLGSEMEDKAGLTDEQ- IADNITIGVFAARDY 279
 QY 305 TASAASTSLTYLLGYPHVLOKVRBELSKGLCKSNQDNKIDMELIQKIGYIKETL 364
 DB 280 TASVLTWILKLYLDNPVLAAVLEE--QMAIRDKKKGESLJWEDTKMPLTYRVIOETL 337
 QY 365 RLNPVPGGEFVALKTEPLNGYOIPKGMNIVYSICDTHDVAEIFTNKEEFNDRFS-APH 423
 DB 338 RAATILSFTRAEVDEVEYEGYILIPKGMKVLPLFRNIHNAADIFSDGKDPDSFEVAPK 397
 QY 424 PEDASRFSFIPEGGLRSCYCKGEPAKILLKIFVVELARHCWQLLNGPPTKTSPTVYV 483
 DB 398 PN-----TFMFGSGHSCGNELAKLEISVLJHLHTTKYRWSIVGSDGIQYGFALPQ 452
 QY 484 DNP 487
 DB 453 NGLP 456

RESULT 11
 O65624 PRELIMINARY; PRT; 457 AA.
 AC O65624;
 DT 01-AUG-1998 (TREMBLrel. 07, Created)
 DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE CYTOCHROME P450.
 GN T18B16.200 OR ATAG19230.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 OC Brassicales; Brassicaceae; Arabidopsis.
 NCBI_TaxID=3702;
 RN (1)
 RP SEQUENCE FROM N.A.
 RA Bevan M., Benes V., Reclmann S., Borkova D., Ansoorge W., Bancroft I.,
 RA Mewes H.W., Mayer K., Schueller C.,
 RA Submitted (JAN-1998) to the EMBL/Genbank/DBJ databases.
 RN (2)
 RP SEQUENCE FROM N.A.
 RA Benes V., Reclmann S., Borkova D., Ansoorge W., Mewes H.W., Lemcke K.,
 RA Mayer K.F.X.,
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
 RN (3)
 RP SEQUENCE OF 131-457 FROM N.A.
 RA Van Der Schueren J., Chuang Y.J., Aert R., Defoor E., Robben J.,
 RA Volckaert G., Mewes H.W., Lemcke K., Mayer K.F.X.,
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.

[4]
 RN SEQUENCE FROM N.A.
 RA Arabidopsis sequencing project;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 DR EMBL; AL021687; CAI6713.1; -;
 DR EMBL; AL161550; CAB78925.1; -;
 DR HSSP; P33006; ICP1.
 DR InterPro; IPR001128; -;
 DR Pfam; PF00067; P450; 1.
 DR PRINTS; PR00385; P450.
 DR PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_1.
 DR Heme; Monooxygenase; Oxidoreductase.
 KW SEQUENCE 457 AA; 52436 MW; 6105FB/C181E4F07 CRC64;
 SQ
 Query Match 19.6%; Score 512; DB 10; Length 457;
 Best Local Similarity 29.6%; Pred. No. 1.3e-33;
 Matches 144; Conservative 84; Mismatches 204; Indels 54; Gaps 13;
 QY 18 LLLFLAATKMDLYCVSGRDR--SCALPLPGTGMGFFPFGFTTQMVLQ-RRKFLQMKRRK 74
 DB 8 LTLFAGSLFLVFLMLQLISQRRFGSSKLPLPGTGMGPVYGETFQLYSODPVVFFQSKQR 67
 QY 75 YGFIYKTHLFGRPYRWGADNVRRIILGDRLYSVHMPASVFTILSGCLSNLHDSHK 134
 DB 68 YGSVEKTHVLCPCVMISPEAKFVLTKSHLKPFPASKERMLGQALFFHQGDYHA 127
 QY 135 QRRKTYMAFSEALCEYVPVITEVSGSLBQWLSCEGRLV--YPEKRIAMRIARI 192
 DB 128 KIRKLYLAFNPESIRNANVPDIESIADSLASW-----EGMINTYQMKTYTNNALLS 182
 QY 193 LLGCPQLAGDSEQQLVEAFEEKTRNLFSLPIDVPSGLRGKAKARLIHARIEONIR 252
 DB 183 IFGKPEVL-----YREDLRCGYILKEKYNMPVNLPGTLFHKSKAKREL-----SOIL 232
 QY 253 AKICGLRASEAGGCKDALQILIEHSEMERGLDQALKOSTELLEFGHETTASATSL 312
 DB 233 ARI-----LSERRONGSSHNLDLGSFGDKBELTDBQ-IADNIGVIFAAROTITASVMSWI 287
 QY 313 IYVLEPHVLOKVEELKSKGLCKSNQDNKLMELIEQYIGCVIKETLRILNPPVPG 372
 DB 288 LKYLLENPNVLEAVTEE--QMAIRKDKREGESLTMGDKRKKMPLSRVIGETLRVASILSF 345
 QY 373 GFRVALKTFELNGYQIPKGMNVIYSICDTHVAELFTKKEPNPREFS-AHPPEASRFS 431
 DB 346 TRREAVVEVEYEGYILPRGMKVLPLFRNIHNSADIFSNPGKDPSEFVAPKPN-----T 400
 QY 432 FLRPGGRLSCVSGKEFAKILKTI---FTVELARHCDQLNGPPTMTKSPVYVVDNLP 487
 DB 401 FMPFGNGHSCGNEKLAKEMSIHHLITTKRCVCVYLL-----IT 443
 QY 488 ARFTTF 493
 DB 444 FSETHF 449
 RESULT 12
 ID 081077 PRELIMINARY; PRT; 482 AA.
 AC 081077;
 DT 01-NOV-1998 (TREMBLrel. 08, Created)
 DT 01-NOV-1998 (TREMBLrel. 08, last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, last annotation update)
 DE PUTATIVE CYTOCHROME P450.
 GN T914.17.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 OC Brassicales; Brassicaceae; Arabidopsis.
 OC NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN-CV. COLUMBIA;
 RA Rounsley S.D., Lin X., Kaul S., Shea T.P., Fujii C.Y., Mason T.M.,
 RA Shen M., Rensing C.M., Fraser C.M., Somerville C.R., Venter J.C.;
 RA *Arabidopsis thaliana chromosome II BAC T914 genomic sequence.
 RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 DR EMBL; AC005315; AAC33235.1; -;
 DR InterPro; IPR001128; -;
 DR Pfam; PF00067; P450; 1.
 DR PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_1.
 DR Heme; Monooxygenase; Oxidoreductase.
 KW SEQUENCE 482 AA; 55175 MW; AB6A07AB2778DD3B CRC64;
 SQ
 Query Match 18.9%; Score 492.5; DB 10; Length 482;
 Best Local Similarity 28.4%; Pred. No. 5.5e-32;
 Matches 141; Conservative 90; Mismatches 222; Indels 43; Gaps 12;
 QY 4 PALLASALCFVPLBLFLAATKMDLYCVSGRDRSCALPLPGTGMGFFPFGFTTQMVLQ 63
 DB 20 PALTLTIVVVVLLF-----KWLHMKQR-----LRPGSGMLPYIGETLRILYTE 68
 QY 64 R-RKFLQMKRRKYGFYKTHLFGRPYRWGADNVRRIILGDRLYSVHMPASVFTILGS 122
 DB 69 NPNSEFATRQNTYGDIFKTHILGCPVMISPEARAVLYSKAHLKPTYPPESEKRMIGP 128
 QY 123 GCISNLHDSHKKQKAYIMAFSEALCEYVPVITEVSGSLBQWLSCEGRLVPEVK 182
 DB 129 EALFFHGGPASHLTKRLVQSSFWPSALRPTVSHIELIVLTLSMTSOKSINTLEY--MK 186
 QY 183 RLMEFRIARILGCEPOLAGDSEQD-----LVEAFEEKTRNLFSLPIDVPSGLRGK 237
 DB 187 RYAFDVAI-----MSAFQKKEPTTIDYIKLYQRLERGNYSMPDLPTLFRKSW 237
 QY 238 KARNLHARIEONIRAKICGLRASEAGGCKDALQILIEHSEMERGLDQALKOSTEL 297
 DB 238 KAR-----TEISEELRKVIEKRENGREEG--GLGVLGAKDKRNGLSQSIADNIGV 291
 QY 298 LFGHETTASATSLIYVLYGLYPHVLOKV--REELKSKGLCKSNQDNKLMELIEQKYL 356
 DB 292 IFAATDTASVLTWLYLKHDPVLDQEVSRBQPSINOKIKKER--RIMWETRKMPLR 349
 QY 357 GCVIKETLRILNPPVGGFRVALKTFELNGYQIPKGMNVIYSICDTHVAELFTKKEPNP 416
 DB 350 TRVIGETLRVASVLSFPRFRAVQVDEYDGLIPRGMKVLPLFRNIHNSSEFPDPEKEDP 409
 QY 417 DRFS-AHPPEASRFSFLRPGGRLSCVSGKEFAKILKTIFTVELARHCDQLNGPPTMK 475
 DB 410 SRFEVAPKPN-----YTPMPFGNGVHSCGSELALEMILHLITTSFRWEVIGDEGIG 464
 QY 476 TSPVYVVDNLPART 491
 DB 465 YGPPVVRKGLPIRYT 480
 RESULT 13
 ID 09FMA5 PRELIMINARY; PRT; 465 AA.
 AC 09FMA5;
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, last annotation update)
 DE CYTOCHROME P450.
 GN Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 OC Brassicales; Brassicaceae; Arabidopsis.
 OC NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-COLUMBIA;
 RX MEDLINE=96290546; PubMed=9628582;
 Sato S., Kaneko T., Kotani H., Nakamura Y., Asamizu E., Miyaajima N.,

RA Tabata S.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. IV.
 RT Sequence features of the regions of 1,456,315 bp covered by nineteen
 RT physically assigned P1 and PAC clones."
 RL DNA Res. 5:41-54(1998).
 DR EMBL; AB009048; BAB08653.1;
 SQ SEQUENCE 465 AA; 53767 MW; 0C00459C8C66D1F CRC64;

Query Match 17.3%; Score 450.5; DB 10; Length 465;
 Best Local Similarity 27.8%; Pred. No. 1.4e-28;
 Matches 136; Conservative 96; Mismatches 203; Indels 55; Gaps 15;

QY 1 MGRLPALLASALCTFVLPFLLEFLAIAIKLMD--LYCVSGRDSICALPLPGTWGPFPGFTL 58
 DB 8 MGLLLIIVS-LC-----SALLRMNOMRYTKNG-----LPPGIMGWPIFGETT 48
 QY 59 QMVLQRRKFLQMKRRKGYFTKTHLFGRPVRYMGADNVRILLGDDRLYSVHPASVRT 118
 DB 49 EFLKQGFNFMNRQRLRYGSPFKSHLGCPTLLISMDSVNRYLKNEKGLVPGYPQSMLD 108
 QY 119 ILTSGCLSNLHSDSHKORKKIVIRAFSREALCEY-VPIIEVGSSLEQMLSCGERGLL- 176
 DB 109 ILTQCNMAAVHGSSSHRLMRSLSLISSTMARDHILPKVDHFMRSYLDOW---NELEVID 165
 QY 177 VYPEVKRLMFRIMRILLCGCEPOLAGDGDSEQLVE---AFEEMTNLFSLPIDVPSFG 232
 DB 166 IQDKTKHMAF-----LSSLTQIA--GNLRKPEVEEFKAFELVYGTLSVPIDLEGTN 216
 QY 233 LYGMKARNILIHARIEONIRAKICGLRASEAGGCKDALOLLIEHSEMERELDMQALKQ 292
 DB 217 YRCGIRARNRIDILRLIMQER-----RDSGETFTDMLGYLMKEGNGRYPLTD--EIRID 269
 QY 293 SSTLELFGHETASATSLITYLGLYPHYLOKREELSKGLLCKNOONKIDMETLEQ 352
 DB 270 QVYTIILSGEYVSTISMAALKYLHDPKALQELRAE--HLAREKRODPELGLDYKVS 327
 QY 353 LKYGICVKEFTLRLNPVPGGFVALKTEFLNGYQIPKGNVYISICDTHDAEIFTNKE 412
 DB 328 MKETRAVIVYETSLATIVNGVLKRTTRDELINGLIPKGRIVYYTREINVDANLYEDPL 387
 QY 413 ENPDRFSAHPEDASRFSPFGGGLRSCVCGKEFAIKLITFVELARCDMQLNGPP 472
 DB 388 IFPMRMKMKSLD--SONSCFVGCGTRLCPEKELGIVEISSFLHYFVTRYRWEIGDE 445
 QY 473 TMTKSTVYP 482
 DB 446 LM-----VFP 450

RESULT 14
 ID 064989 PRELIMINARY; PRT; 513 AA.
 AC 064989;
 DT 01-AUG-1998 (TREMblrel. 07, Created)
 DT 01-AUG-1998 (TREMblrel. 07, last sequence update)
 DE 01-MAR-2001 (TREMblrel. 16, last annotation update)
 DE STEROID 22-ALPHA-HYDROXYLASE.
 GN DMF4.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 OC Brassicales; Brassicaceae; Arabidopsis.
 OC NCBL_TaxID=3702;
 OX [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=MS-2;
 RA Choe S., Dilkes B.P., Fujioaka S., Takatsuto S., Sakurai A.,
 RA Feldmann K.A.;
 RL Plant Cell 10:0-0(1998).
 RL -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 DR EMBL; AF044216; AAC05093.1; -
 DR InterPro; IPR001128; -.

DR Pfam; PF00067; P450.1.
 DR PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN.1.
 KW Heme; Monooxygenase; Oxidoreductase.
 SQ SEQUENCE 513 AA; 58867 MW; B1639BDD9A4DA6F3 CRC64;

Query Match 16.8%; Score 439; DB 10; Length 513;
 Best Local Similarity 27.0%; Pred. No. 1.4e-27;
 Matches 142; Conservative 85; Mismatches 226; Indels 72; Gaps 15;

QY 3 LPALLASALCTFVLPFLLEFLAIAIKLMDLYCVSGRDSICALPLPGTWGPFPGFTL---- 58
 DB 10 LPILLPSL-----LSSLFLILK-----RRNRKTRFNLPPGKSGWPIFGETTGYLK 57
 QY 59 QMVLQRRKFLQMKRRKGYFTKTHLFGRPVRYMGADNVRILLGDDRLYSVHPASVRT 117
 DB 58 PYTATTLGDEWQOQHVSKYKGYRSLNLFCEPPIVSADAGLNFILLQNGRLFECSYPSISG 117
 QY 118 ILTSGCLSNLHSDSHKORKKIVIRAFSREALCEYVPIIEVGSS-----LEQMLSCGER 173
 DB 118 GILCKSMILVYGDMDMRSLNFISSHARL---TILKDVRRHLLFVDSMQO--NS 172
 QY 174 GLVYPEVKRLMFRIMRILLCGCEPOLAGDGDSE--QQLVEAFEEMTNLFSLPIDVPSFG 232
 DB 173 IFSADBAKFKFTFMIAKHMISMDP-----GEETEDLKKEYVTFMKGVYSAPIPLPGTA 227
 QY 233 LYRMKARNILIHARIEONIRAKICGLRASEAGGCKDALOLLIEHSEMERELDMQALKQ 292
 DB 228 YHKALQSRATILKFEKMEERKIDKEED---QEEEVKTDEAEMSKSDHVRKORT 282
 QY 286 DMQ-----ALKQS--STE-----LFGHETASATSLITYLGLYPHYLOKREELK 331
 DB 283 DDDELGVWLKHSNSTQIIDLILSLFAGHETSSVAIALAIFLQACPAVEELREHL 342
 QY 332 SKGLLCSNODNKIDMETLEQKTCYVKEFTLRLNPVPGGFVALKTEFLNGYQIPKG 391
 DB 343 ELARAKKELESESLMWDYKMDPTQCVINETRLGVVYRLHRAKLDVRYKXYDIPSG 402
 QY 392 WNVYISICDTHDAEIFTNKEENPDRFSAHPEDASRF-----FIPFGGLRSC 442
 DB 403 WKVLPVISAHLNDSRKDQRLNLPFRWQOONNGASSSGSFSTWNNNTMFPGGGRFLC 462
 QY 443 VKKEPAKILKIFTELARHCDMQLNGPPMTKTSPTVYVDNLP 487
 DB 463 AGSELAKLEMAVFTHLVLKFNWELAE-----DDQPPAPFVDP 502

RESULT 15
 ID 09SCQ9 PRELIMINARY; PRT; 513 AA.
 AC 09SCQ9;
 DT 01-MAY-2000 (TREMblrel. 13, Created)
 DT 01-MAY-2000 (TREMblrel. 13, last sequence update)
 DE 01-MAR-2001 (TREMblrel. 16, last annotation update)
 DE STEROID 22-ALPHA-HYDROXYLASE (DMF4).
 GN T3A5.40.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 OC Brassicales; Brassicaceae; Arabidopsis.
 OC NCBL_TaxID=3702;
 OX [1]
 RN SEQUENCE FROM N.A.
 RA Bioecker H., Mewes H.W., Lemcke K., Mayer K.F.X., Quetier F.,
 RA Salanoubat M.;
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 RN [12]
 RP SEQUENCE FROM N.A.
 RP EU Arabidopsis sequencing project;
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
 RL -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 DR EMBL; AL132979; CAB62435.1; -
 DR InterPro; IPR001128; -.

DR Pfam: PF00067; P450; 1.
DR PRINTS: PR00385; P450.
DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
KW Heme; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 513 AA; 58867 MW; B1639BDD9A5D7C93 CRC64;

Query Match 16.8%; Score 439; DB 10; Length 513;
Best local similarity 27.0%; Pred. No. 1,4e-27;
Matches 142; Conservative 85; Mismatches 226; Indels 72; Gaps 15;

```
QY 3 LPLLALSACTFVLPILLPLIAAKIMDLICVSGRDSICALPLPGTMGFPFGETL-----58
   |||  :  :  |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 10 LPLLPLPSL-----LSLPLFLILK-----RRNRKTRFNLPFGSKSGWPFGLGTIGYLK 57

QY 59 -QWVLDORRKFLOMKRRRYGFIYKTHLFGPRTVVMGADNVNRILLGDDRLVSVHMPASVR 117
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 58 PYATATIGDPMQOHVSKYGIYRSNLFGEPTIYSADAGLNRFILQNEGRLFECSYPRSIG 117

QY 118 TILGSGCLSNLHDSHKKRKKVIMRAFSREALCYVPVITEVGS-----LEQWLSCGER 173
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 118 GILGKMSMLVLDGDMHDMRSISLNFISHARLR---TILKDYERHTLFEVLDSWQO--NS 172

QY 174 GLVYPPVKKRMFRIMRILLGCEPOLAGDSE-QOLVFAFEEMTRNLFSLPIDVPFSG 232
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 173 IFSADDEAKKFTFNLMAKHIMSMDP-----GEEETEQLKKEIYTFMKGVYSAFLNLPFTA 227

QY 233 LYRGMKARNLIHARIEQNIIRAKICGLRASEAGOGCKDALQLIEHSWE-----RGERL 285
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 228 YHKALQSRATILKFIERKMERKIDIKEED-----QEEEVKTEDEAEMSKSDHVRKQRT 282

QY 286 DMQ---ALKOS--STE-----LLFGHETASATSLITYLGLYPHYLOKVRRELK 331
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 283 DDDLLGWVLKHSNLSSTEOILLILSLFAGHETSSVAIALAIEFLQACPRAVEELREHL 342

QY 332 SKGLLCKSNQDKLDMELDLQKLYIGCVIKETLRINPVPGPFRVALKTFELNGYQIPKG 391
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 343 EIAKAKKELGSEELNMDYKKMDFTQCYNETLRKGNVRFIHRKALKDVRKYQYDIPSG 402

QY 392 WNVLYSICDTHDVAIEFTNKEEFNDRFSAPHPEDASRFS-----FIPGGGLRSC 442
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 403 WKVLPIYSIVHLDSRYDQPLFNPWRMOQNNGASSSGSFSFTWGNMYMPFGGPRLC 462

QY 443 VGRFEAKILKIFVELLARHCDQMLNGPPTMKTSPTIYVPVDNL 487
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 463 AGSELAKLEMAVFTIHLVLKFNWELAE-----DDKPAFPVPVDFP 502
```

Search completed: November 6, 2001, 13:38:27
Job time: 255 sec


```

139 GCGCGCGCGACCGCACTTGTCCTCCATTTGCCCGCCCGGAGCTATGGCG 188
51 pheProPhephegLygluThrLeuGlnMetValLeuGlnArgArgLysPhe 67
189 TTCCTCTTTTGGGAAACCTTGCAATGATGTACTGACGGGAGGAGATT 238
67 eLeuGlnMetLysArgArgLysTyrGlyPheIleTyrLysThrHisLeuP 84
239 CCTGCAGATGAAGCGCAGGAATAACGGCTTCATCTACAAGACCATCTGT 288
84 heGlyArgProThrValArgValMetGlyAlaAspAsnValArgValIle 100
289 TCGGGGGGCCACCGCTACGGGTGATGGCGCGCAATGTGCGGCCCATC 338
101 LeuLeuGlyAspAspArgLeuValSerValHisTrpProAlaSerValAr 117
339 TTGCTCGGAGACGACCGGCTGTGTGTGCTCCACTGGCCAGCGTGGTGG 388
117 gThrIleLeuGlySerGlyCysLeuSerAsnLeuHisAspSerSerHisL 134
389 CACCATTTGGGATCTGGCTGCTCTTAACCTGCACGACTCCTGCACAA 438
134 yGlnLArgLysLysValIleMetArgAlaPheSerArgGluAlaLeuGlu 150
439 ACGACGCGAAGAGGTGATTATGGGGCCTTCAGCCGCGAGCGACTCGAA 488
151 CysTyrValProValIleThrGluGluValGlySerSerLeuGlnLysTr 167
489 TCCTACGTCCGGTGTATCACGAGGAAGTGGCGACAGCGCTGGACGAGT 538
167 pLeuSerCysGlyGluArgGlyLeuLeuValTyrProGluValLysArgL 184
539 GGTGAGCTGGCGGAGCGCGGCTCTCTGCTTAACCCGAGGTGAACCGCC 588
184 eumePheArgIleAlaMetArgIleLeuLeuGlyCysGluProGlnLeu 200
589 TCATGTTCCGAAATCCGCATGCCCATCTACTGGGCTGGACACCCCAACT 638
201 AlaGlyAspGlyAspSerGluGlnGlnLeuValGluAlaPheGlnLys 217
639 GCGGGCGACGGGAGCTCCGACGACAGCTGTGTGAGGCGCTTGAGAGAA 688
217 tThrArgAsnLeuPheSerLeuProIleAspValProPheSerGlyLeuT 234
689 GACCCGCATCTCTTCTGCTGCCATCCATCGACGTGCCCTTCACCGGGCT 738
234 yTArgGlyMetLysAlaArgAsnLeuIleHisAlaArgIleGlnLysAs 250
739 ACCGGGGCATGAAGCGCGGAACTCATTCACGCCGCGCATCGACGAGAAC 788
251 IleArgAlaLysIleCysGlyLeuArgAlaSerGluAlaGlyGlnLysC 267
789 ATTCCGGCCAAAGATCTGGCGGCTGGCGGATCCGAGGCGGCGCAAGGCT 838
267 sLysAspAlaLeuGlnLeuLeuIleGlnHisSerTrpGluArgGlyGlu 284
839 CAAGACGCGCTGCGAGCTGTGATCGAGACCTCGGGGAGAGGAGGAGAC 888
284 rGLeuAspMetGlnAlaLeuLysGlnSerSerThrGluLeuLeuPheGly 300
889 GCGTGCAGATGAGGAGCTAAAGCAATCTTCAACCGAATCCTCTTTGGA 938
301 GlyHisGluThrThrAlaSerAlaAlaThrSerLeuIleThrTyrLeuG 317
939 GGACACGAAACCGCGCAGTGCACGCCACATCTGTGATCACTTACTTGG 988
317 yLeuTyrProHisValLeuGlnLysValArgGluLeuLeuLysSerLys 334
989 GGTCTACCCACATGTTCTCCAGAAAGTGCAGAAAGAGCTGAAGAGTAAG 1038
334 yLeuLeuCysLysSerAsnGlnAspAsnLysLeuAspMetGluIleLeu 350
1039 GTTACTTGTGCAAGAGCAATCAAGACAAAGTTGAGACATGAAATTTTG 1088

```

```

351 GluGlnLeuLysTyrIleGlyCysValIleLysGluThrLeuArgLeuAs 367
1089 GAAACAATTAAATACATCGGGTGTGTTATTAAAGACAGACCTTGACGTGA 1138
367 nProProValProGlyGlyPheArgValAlaLeuLysThrPheGluLeuA 384
1139 TCCCCAGATTCCAGAGAGGGTTTGGGTGTGCTCGAAGACTTTGAATTAA 1188
384 snGlyTyrGlnIleProLysGlyTyrAsnValIleTyrSerIleCysAsp 400
1189 ATGAGTACCAAGATTCACAAAGGCTCGAATGTTATCTACAGTATCTGTAT 1238
401 ThrHisAspValAlaGluIlePheThrAsnLysGluGlnPheAsnProAs 417
1239 ACTCATGATGTGGCAGAGATCTTCACCAACAGAGAAAGAAATTATCTGTA 1288
417 pArgPheSerAlaProHisProGluAspAlaSerArgPheSerPheIleP 434
1289 CCGATTTCATGCTGCTCACCCACAGAGATGCATCCAGTTCCAGCTTCATTC 1338
434 ropPheGlyGlyGlyLeuArgSerCysValGlyLysGluPheAlaLysIle 450
1339 CATTTGAGAGAGGCTCTAGAGCTGTAGGCAAAAGAAATTTGCCAAAATTT 1388
451 LeuLeuLysIlePheThrValGluLeuAlaArgHisCysAspTrpGlnLe 467
1389 CTTCCTAAATATATTACATGGAGCTGGCCAGGCAATTCAGCTGACAGCT 1438
467 uLeuAsnGlyProProThrMetLysThrSerProThrValTyrProValA 484
1439 TCATAATGTGACCTCTTACATGAATAAACAGTCCACCGTATCTGTGG 1488
484 sPAsnLeuProAlaArgPheThrHisPheHisGlyGluIle 497
1489 ACAATCTCCCTGCAAGATTACCCATTTCCATGGGGAATTC 1529

seq_name: gb_tol:MMP450RA
seq_documentation_block: 1756 bp mRNA ROD 08-JAN-1998
LOCUS MMP450RA
DEFINITION M.musculus mRNA for P450RA protein.
ACCESSION Y12657
VERSION Y12657.1 GI:2765213
KEYWORDS P450RA gene.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE
1 (bases 1 to 1756)
FujiI,H., Sato,T., Kaneko,S., Gotoh,O., FujiI-Kuriyama,Y.,
Osawa,K., Kato,S., and Hamada,H.
Metabolic inactivation of retinoic acid by a novel P450
differentially expressed in developing mouse embryos
TITLE
JOURNAL
REFERENCE
AUTHORS
TITILE
JOURNAL
FEATURES
SOURCE
1. 1756
/organism="Mus musculus"
/strain="C3H"
/db_xref="taxon:10090"
/cell_line="P19"
57..1550
/gene="P450RA"
/gene="P450RA"
57..1550
/function="metabolises retinoic acid"
/codon_start=1
/product="P450RA protein"

```

/protein_id="CAA73206.1"
/db_xref="GI:2765214"
/translation="MGLPALLASALCTFVLPLELLFLAALKIMLDYCVSSRDSCALPL
PGPMGFPEFETLQVLORRKFLQMLDSSKQKRYLYKTHLFRPTVRVGMADNVRIL
GEHLVSVHMPASVRLTLAGCLSNLHSSKQKRYLYKTHLFRPTVRVGMADNVRIL
SSCLQWLSCGERGLYPEVKRLMFRIMRILGCEPGRAGGEDDQQLVEAREMT
RNLFSLPIDVPSGLYRGVKARNLIHAETENIAAKTIRLQATEPDGCKDALDLLE
HSMRGERLDMQALKOSTELLFGHEHTAASLITVGLYPHVLQKARETSKG
LLCKSNODKLDMETLEQLKYIGCVIKETTLRLPNVPGFVALKTFELNGYQIPKM
NVITSCDTHVDVIFTKKEFNDRFIVPEASRSRSLPFGGLRSQVGRFAKI
LKTFTVLAHRCOMQLNGRPMTKTSPTVYPVNLPRAFYFGDL"

BASE COUNT 408 a 434 c 494 g 420 t
ORIGIN

alignment_scores:
Quality: 2437.00 Length: 497
Ratio: 5.035 Gaps: 0
Percent Similarity: 97.384 Percent Identity: 93.360

alignment_block:
US-09-668-482-4 x MMP450RA ..

Align seg 1/1 to: MMP450RA from: 1 to: 1756

1 MetGlyLeuProAlaLeuAlaLeuAlaSerAlaLeuCysThrPheValLeuPr 17
|||||
57 ATGGGGCTCCGGCGCTGCTGGCCAGTGGCTGACCTTGCCTGCTGCC 106
|||||
17 OLeuLeuPheLeuAlaAlaIleLysLeuThrPheLeuTyrCysValS 34
|||||
107 GCTGCTGCTTCTTCTGGCGGCTCAAGCTCTGGACCTGTACTGCTGTA 156
|||||
34 ergIYATGASPARSerCysAlaLeuProLeuProGlyThrMetGly 50
|||||
157 GCACCGGAGTGCAGCTGCCCTCCCTTGGCCCCCGGATACCATGGCC 206
|||||
51 PheProPhePheGlyIuThrLeuGlnMetValLeuGlnArgArgLysPh 67
|||||
207 TTCACATCTTGGGGAACATGTCAGATGCTGCTCAGCGAGAGATT 256
|||||
67 eLeuGlnMetLysArgArgLysTyrGlyPheIleTyrLysThrHisLeu 84
|||||
257 TCTGACAGTGAAGCGCAGAAATACGGCTTCATCTACAGACGATCTGT 306
|||||
84 heGlyArgProThrValArgValMetGlyAlaAspAsnValArgArgIle 100
|||||
307 TTGGCGGCGCCACGCTGGGCTGATGGCGCGGATATGTGGCGGATC 356
|||||
101 LeuLeuGlyAspAspArgLeuValSerValHisThrProAlaSerValAr 117
|||||
357 TTGCTGGGAGAGACCGGTTGCTGCTGCATCGCCCGCGTGGTGG 406
|||||
117 gThrIleLeuGlySerGlyLysLeuSerAsnLeuHisAspSerHisL 134
|||||
407 CACCATCTGGGCGCTGGCTGCTCCACCTGCACGATTCCTCGCACA 456
|||||
134 ySglnArgLysLysValIleMetArgAlaPheSerArgLysAlaLeuGln 150
|||||
457 AGCAGCGAAGAAGGTGATTAAGCGCTTCAGCGCGAGGACACTCCAG 506
|||||
151 CysTyrValProValIleThrGlnGlnValGlySerSerLeuGlnGlnTr 167
|||||
507 TCTCTACGTGCTGATCGCTGAGAGACGACAGTGTCTGCGACAGTGG 556
|||||
167 pheuSerCysGlyLysArgGlyLeuLeuValTyrProGlnValLysArgL 184
|||||
557 GCTAAGCTCGGCGAGCGGCGCTCTGCTACCCGAGGTGAAGCGCC 606
|||||
184 eumetPheArgIleAlaMetArgIleLeuLeuGlyCysGlnProGlnLeu 200
|||||
607 TCATGTTCCGATCGCATGCGATCTGCTGGCTGGAGCGGGGTCCA 656

201 AlaGlyAspGlyAspSerGlnGlnGlnLeuValGlnAlaPheGlnGlnL 217
|||||
657 GGGGGCGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAT 706
|||||
217 tThrArgAsnLeuPheSerLeuProIleAspValProPheSerGlyLeuT 234
|||||
707 GACCCGCAATCTCTCTCTCTCCCATGAGCTGCTTATGAGGCGCTGT 756
|||||
234 yArgGlyMetLysAlaArgAsnLeuIleHisAlaArgIleGlnGlnAsn 250
|||||
757 ACCGGGCGGTGAAGCGCGGAGACCTTATACAGCCGATCGAGAGAAC 806
|||||
251 IleArgAlaLysIleCysGlyLeuArgAlaSerGlnAlaGlyGlnGlyC 267
|||||
807 ATTCGGCCAGATCCGCGGCTTACAGCTACAGACCGGAGGGGGGTGG 856
|||||
267 sLysAspAlaLeuGlnLeuLeuIleGlnHisSerTrpGlnArgLysL 284
|||||
857 CAAGGAGCGGCTGCAGCTCTGATTAAGCAGCTGAGGAGGAGAGAGA 906
|||||
284 rGleuAspMetGlnAlaLeuLysGlnSerSerThrGlnLeuLeuPheGly 300
|||||
907 GCGTGGATATGCAAGGCATTAACAATCGTCAACAGAGCTCTCTTGGT 956
|||||
301 GlyHisGlnThrThrAlaSerAlaIleThrSerLeuIleThrTyrLeuG 317
|||||
957 GGTCAATGAACATACAGCAGTGTGCGACATCATCATCTACTTACTAG 1006
|||||
317 yLeuTyrProHisValLeuGlnLysValArgGlnGlnLeuLysSerLysG 334
|||||
1007 ACTCTACCCACATGCTCTCCAGAAAGTTTCAGAAAGATAAAGAGCAAG 1056
|||||
334 yLeuLeuCysLysSerAsnGlnAspAsnLysLeuAspMetGlnLeu 350
|||||
1057 GCTTACTTTGCAAGACCAATCAAGACACAGATTAAGATGGAACCTTGG 1106
|||||
351 GlnGlnLeuLysTyrIleGlyCysValIleLysGlnThrLeuArgLeuAs 367
|||||
1107 GAACAGCTTAATATACATTGGGTGTGTCATTAAAGAGACCTCGATTGAA 1156
|||||
367 nProProValProGlyGlyPheArgValAlaLeuLysThrPheGlnLeu 384
|||||
1157 TCTCCGGCTCCAGAGAGGTTTCGGGTGCTCGAAGACTTTGAGCTGA 1206
|||||
384 snGlyTyrGlnIleProLysGlyTyrAsnValIleTyrSerIleCysAsp 400
|||||
1207 ATGATATCCAGATCCCAAGGCTGGAGTGTATTATACAGTATCTGTGAC 1256
|||||
401 ThrHisAspValAlaGlnIlePheThrAsnLysGlnIlePheAsnProAs 417
|||||
1257 ACCCAGCATGTGGCAGATCTCTACATTAACAGAGAGAAATTAATCCCG 1306
|||||
417 pArgPheSerAlaProHisProGlnAspAlaSerArgPheSerPheIle 434
|||||
1307 CCGCTTATATAGTGCCTCATCCAGAGGATGCTCCCGGTTGAGCTTAATC 1356
|||||
434 ropheGlyGlyGlyLeuArgSerCysValGlyLysGlnPheAlaLysIle 450
|||||
1357 CATTTGAGAGGAGCTTGGAGCTGTGTAGGCAAAAGTTTGCAAAAT 1406
|||||
451 LeuLeuLysIlePheThrValGlnLeuAlaArgHisCysAspTrpGlnLe 467
|||||
1407 CTCTTAAGATATTACAGTGAAGCTGTAGGACACGTGATTTGGCAGCT 1456
|||||
467 uLeuAsnGlyProProThrMetLysThrSerProThrValTyrProVal 484
|||||
1457 TCTAAATGAGCTCTCTCAATGAAGACAAAGCCCATGCTGTACCTGTGG 1506
|||||
484 sPanLeuProAlaArgPheThrHisPheHisGlyGlnLe 497
|||||
1507 ACAATCTCCCGCAAGATTAACCTTCACTTCCAGAGGAGATATC 1547

seq_name: gb_rol:AF115769
seq_documentation_block: 1726 bp mRNA ROD 03-MAR-1999
LOCUS AF115769
DEFINITION Mus musculus cytochrome P450 retinoic acid metabolizing enzyme
P450A mRNA, complete cds.
ACCESSION AF115769
VERSION AF115769.1 GI:4324987
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS Paine,C.T., Paine,M.L. and Snead,M.L.
TITLE 1 (bases 1 to 1726)
JOURNAL Identification of amelogenin- and tufelatin-interacting proteins
using the yeast two-hybrid system
Connect. Tissue Res. 38, 257-267 (1998)
2 (bases 1 to 1726)
AUTHORS Paine,C.T., Paine,M.L. and Snead,M.L.
TITLE Direct Submission
JOURNAL Submitted (21-DEC-1998) CCMB, University of Southern California,
2250 Alcazar Street, GSA103, Los Angeles, CA 90033, USA
FEATURES
SOURCE
1..1726
/organism="Mus musculus"
/db_xref="taxon:10090"
/tissue_type="tooth"
/dev_stage="1 day old"
27..1520
/note="tufelatin-interacting protein 24 TTP-24"
/codon_start=1
/product="cytochrome P450 retinoic acid metabolizing
enzyme P450A"
/protein_id="AA017217.1"
/db_xref="GI:4324987"
/translation="MGLPALATATLCTFVLPLFLAIAKIMLYCYSSRRSCALPL
PGMGPPFGGCTQWTLQRRKFLQMKRRKGLFYKTHLPERPYRVKGNVRIIL
GERRHYVNPASVRTILGAGCUNLHDSHKOKKVMQFSEALQCYLVLAEEV
SSCLEQWLSCGERLIVPEVKRIAMRIILGCEPAGSGEDBQALVEAEEMT
RNLESLIDVPSGLYGVAKARNLHARIEENIRAKIRLQATPPDGCKALDLLE
HSMWGERLDMOALKOSTELLFGHETASAASTLYGLYHVLQKVEEIKSKG
LICKSNDKLMETLEOLKYGIVKETRLINPVPAGFPVALKTELENGYQIPKGM
NVYISGDTHDVADIPNKKEFNDRIRIVHPEASRPSFLPFGGGLRSCVGRPAKI
LKTFTYELARHCDQMLNGPIMKTSPTYIPVDNLPAKFTYFGDI"
BASE COUNT 405 a 428 c 476 g 417 t
ORIGIN
alignment_scores:
Quality: 2434.00 Length: 497
Ratio: 5.029 Gaps: 0
Percent Similarity: 97.384 Percent Identity: 93.159
alignment_block:
US-09-668-482-4 x AF115769 ..
Align seg 1/1 to: AF115769 from: 1 to: 1726
1 MetGlyLeuProAlaLeuLeuAlaSerAlaLeuCysThrPheValLeuPr 17
|||||
27 ATGGGGCTCCGGCGCTGTGGCACTGGCGCTGTGACCTTGTGTGGTGC 76
17 OleuLeuLeuPheLeuAlaAlaIleLysLeuTrpAspLeuTyrCysValS 34
|||||
77 GCTGCTGCTCTTCCTGCGGCGCTCAAGCTGTGGAGCCGTACTGTGTGA 126
34 exGlyArAspArgSerCysAlaLeuProLeuProGlyThrMetGly 50
|||||
127 GCAAGCCGAGATGCAGCTGCGCTTCCCTCCGCCGCGGTACCAATGGGC 176
51 PheProPhePheGlyGlyThrLeuGlnMetValLeuGlnArgArgLysPh 67
|||||

177 TTCCATTCTTTGGGAAACATTCAGATGCTGCTTCAGCGAGGAAGTT 226
67 eLeuGlnMetLysArgArgLysTyrGlyPheLleTyrLysThrHisLeuP 84
|||||
227 TCTGCAGATGAAGCGAGCAATATGCGCTTCATCTACAAACCATCTGT 276
84 heGlyArProThrValArgValMetGlyAlaAspAsnValArgArgL 100
|||||
277 TTGGGGGCCCCACGCGTGGGTATGGGCGCGGATATATGCGCGGCATC 326
101 LeuLeuGlyAspAspArgLeuValSerValHisTrpProAlaSerValAr 117
|||||
327 TTTCCTGGAGAGCACCGGTTGCTGTGCTGACCTGCGCCGCGGTGGCG 376
117 gThrIleLeuGlySerGlyCysLeuSerAsnLeuHisAspSerSerHisL 134
|||||
377 CACCATCTGGGCGCTGGCTGCTCTCCAACTTCACAGATTCTTCGCACA 426
134 ysgLnaArgLysLysValIleMetArgAlaPheSerArgLysValLeuGlu 150
|||||
427 AGCAGCGAAAGAGTGATATGACAGCGCTTCAGCCGCGAGCACTCCAG 476
151 CysTyrValProValIleThrGluGluValGlySerSerLeuGluGlnTr 167
|||||
477 TGCTACGTGCTGCTGATCGCTGAGGAAGTCAGCGTTGCTGGAGAGTg 526
167 PleuSerCysGlyGluArgGlyLeuLeuValTyrProGluValValArgL 184
|||||
527 GCTAACCTGGCGGAGCGCGGCTCCTGCTACCCGAGGTAAACGCCG 576
184 eumetheArgIleAlaMetArgIleLeuLeuGlyCysGluProGlnLeu 200
|||||
577 TCATGTTCCGATCGCATGCGCATCGCTGCGGTGCGGCGGGTCCA 626
201 AlaGlyAspGlyAspSerGluGlnGlnLeuValGluAlaPheGluGlu 217
|||||
627 GCGGCGCGCGGAGAGCAGACAGCAAGCTCGTAGAGCTTTCAGAGAGAT 676
217 tThrArgAsnLeuPheSerLeuProIleAspValProPheSerGlyLeuT 234
|||||
677 GACCCCAATCTCTCTCTCTCCATTCGACGTCGCTTACGCGGCTG 726
234 yArArgLysMetLysAlaArgAsnLeuIleHisAlaArgIleGluGlnAsn 250
|||||
727 ACCGGGCGCTGAAGGCGGCAACCTTATACGCGGCGCATCGAGGGAAC 776
251 IleArgAlaLysIleCysGlyLeuArgAlaSerGluAlaGlyGlnGlyC 267
|||||
777 ATTGCGGCCAAGATCCGCGCGCTTCAGGCTACAGACCGGATGGGGTGT 826
267 sLysAspAlaLeuGlnLeuLeuIleGluHisSerTrpGluArgGlyGluA 284
|||||
827 CAAGGACGCGCTGCAGCTCCTGATGAGCAGTCGTGGAGAGGAGAGA 876
284 rGluAspMetGlnAlaLeuLysGlnSerSerThrGluLeuLeuPheGly 300
|||||
877 GGCTGATATGCAAGCGCATTAACAATCGTCAACAGAGCTCCTTTGGT 926
301 GlyHisGluThrThrAlaSerAlaIleThrSerLeuIleThrTyrLeuG 317
|||||
927 GGTCAATGAATACAGACCAAGTGTGTCAGACATCATGATATTACTTAG 976
317 yLeuTyrProHisValLeuGlnLysValArgGluGluLeuLysSerLysG 334
|||||
977 ACTCTACCCACATGTCTCTCAGAAAGTTGGAAGAAGAGATGAAGCAAG 1026
334 yLeuLeuCysLysSerAsnGlnAspAsnLysLeuAspMetGluIleLeu 350
|||||
1027 GCTTACTTTGCAAGAGCAATCAAGACAAAGATTAGACATGGAACCTTG 1076
351 GluGlnLeuLysTyrIleGlyCysValIleLysGluThrLeuArgLeuAs 367
|||||
1077 GAACAGCTTAATAATACATTGGGTGTCATTAAGAGAGACCTCGATTGAA 1126


```

367 nProPvalaProGlyGlyPheArVAlaAlaLeuLysThrPheGluLeuA 384
|||||
1127 TCCTCCGCTCCAGAGGGTTTCGGTTCGTCGAGACTTTTGAGCTGA 1176
384 snGlyTyrGluIleProLysGlyTyrAsnValIleTyrSerIleCysAsp 400
|||||
1177 ATGGATRCCGATCCCCCAGGGCTGGATTTATTATACACTATCTGTGAC 1226
401 ThrHisAspValAlaGluIlePheThrAsnLysGluGluPheAsnProAs 417
|||||
1227 ACCACGATGTGCAGATATCTTCACTAACAGAGCAATTTATCCCGA 1276
417 pArgPheSerAlaProHisProLysAspAlaSerArgPheSerPheIleP 434
|||||
1277 CCGCTTTATAGTCCCTCATCCAGAGGATGCTCCCGGTTCAGCTTCATTC 1326
434 ropPheGlyGlyGlyLeuArgSerCysValGlyLysGluPheAlaLysIle 450
|||||
1327 CATTTGAGAGAGGCTTCGAGAGCTGTGTAGCAAAAGATTGCAAAAT 1376
451 LeuLeuLysIlePheThrValGluLeuAlaArgHisCysAspIrrGlnLe 467
|||||
1377 CTTCCTAAGATATTACAGTGGAGCTGTAGGCACATGTGATGGCAGCT 1426
467 uLeuAsnGlyProProThrMetLysThrSerProThrValTyrProValA 484
|||||
1427 TCCTAAATGAGCTCTCTACAAATGAGACACGCCCTGCTGACCTGTGG 1476
484 sPanLeuProAlaArgPheThrHisPheHisGlyIle 497
|||||
1477 ACATCTCCCTCGCAAGATTCACCTACTTCCAGGAGATATTC 1517

```

seq_name: gb_htg21.AL358613

```

seq_documentation_block:
LOCUS      AL358613      166337 bp      DNA           14-APR-2001
DEFINITION Homo sapiens chromosome 10 clone RP11-348J12, *** SEQUENCING IN
PROGRESS ***. 14 unordered pieces.

```

```

ACCESSION  AL358613      GI:13660841
VERSION     AL358613.8
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
ORGANISM    Homo sapiens

```

```

REFERENCE   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS     Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE       1 (bases 1 to 166337)
JOURNAL     Plumb,B.

```

```

COMMENT     Direct Submission
            Submitted (13-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
            On Apr 17, 2001 this sequence version replaced gi:13625010.
            ----- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquerry@sanger.ac.uk
            ----- Project Information
            Center project name: BA348J12
            ----- Summary Statistics
            Sequencing program: XGAP4; version 4.5
            Sequencing vector: plasmid; 108752. 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 161818 bases at least Q40
            Consensus quality: 163957 bases at least Q30
            Consensus quality: 163957 bases at least Q20
            Insert size: 165037; sum-of-contigs
            Insert size: 176845; 5.5% error; agarose-fp
            Quality coverage: 5.93x in Q20 bases; sum-of-contigs quality
            coverage: 5.81x in Q20 bases; agarose-fp
            -----
            * NOTE: This is a 'working draft' sequence. It currently

```

* consists of 14 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

```

1 32409: contig of 32409 bp in length
* 32410 32509: gap of 100 bp
* 32510 52913: contig of 20404 bp in length
* 52914 53013: gap of 100 bp
* 53014 55817: contig of 2804 bp in length
* 55818 55917: gap of 100 bp
* 55918 74512: contig of 18595 bp in length
* 74513 74612: gap of 100 bp
* 74613 83833: contig of 9221 bp in length
* 83834 83933: gap of 100 bp
* 83934 87001: contig of 3068 bp in length
* 87002 87101: gap of 100 bp
* 87102 97488: contig of 10387 bp in length
* 97489 97588: gap of 100 bp
* 97589 100583: contig of 2995 bp in length
* 100584 100683: gap of 100 bp
* 100684 105295: contig of 4612 bp in length
* 105296 105395: gap of 100 bp
* 105396 129702: contig of 24307 bp in length
* 129703 129802: gap of 100 bp
* 129803 132462: contig of 2660 bp in length
* 132463 132562: gap of 100 bp
* 132563 137099: contig of 4537 bp in length
* 137100 137199: gap of 100 bp
* 137200 158309: contig of 21110 bp in length
* 158310 158409: gap of 100 bp
* 158410 166337: contig of 7928 bp in length.

```

FEATURES

source

```

1..166337
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-348J12"
/clone_lib="RPC1-11.2"
1..32409
/note="assembly_fragment:02593"
fragment_chain:1
clone_end:77
vector_side:left"
32510..52913
/note="assembly_fragment:01673"
fragment_chain:1"
53014..55817
/note="assembly_fragment:01466"
fragment_chain:2"
55918..74512
/note="assembly_fragment:01282"
fragment_chain:2"
74613..83833
/note="assembly_fragment:02258"
fragment_chain:2"
83934..87001
/note="assembly_fragment:00938"
fragment_chain:2"
87102..97488
/note="assembly_fragment:01557"
fragment_chain:2"
97589..100583
/note="assembly_fragment:00199"
fragment_chain:2"
100684..105295
/note="assembly_fragment:01309"
105396..129702
/note="assembly_fragment:02028.0"
129803..132462
/note="assembly_fragment:02077"

```

```
misc_feature      132563..137099
                  /note="assembly_fragment:02005
                  fragment_chain:3"
misc_feature      137200..158309
                  /note="assembly_fragment:00282
                  fragment_chain:3"
misc_feature      158410..166337
                  /note="assembly_fragment:01837
                  fragment_chain:3
                  clone_end:SP6
                  vector_side:right"
BASE COUNT      46344 a 33577 c 34990 g 50115 t 1311 others
ORIGIN
```

```
alignment_scores:
  Quality: 2076.50      Length: 1166
  Ratio: 4.221          Gaps: 13
  Percent Similarity: 42.196  Percent Identity: 41.938
```

alignment_block:
US-09-668-482-4 x AL358613 ..

Align seg 1/1 to: AL358613 from: 1 to: 166337

```
1 MetGlyLeuProAlaLeuAlaSerAlaLeuCysThrPheValLeuPr 17
54914 ATGGGGCTCCCGCGCTGCTGGCAGTGGCGCTGACCTGCTGCC 54963
17 OleuLeuLeuPheLeuAlaAlaIleLeuLeuTrpAspLeuTyrCysValS 34
54964 GCTGCTGCTCTTCTGCTGCTGCGATCAAGCTCTGGACCTGACTGCTGA 55013
34 eErGlyArGspArGserCysAlaLeuProLeuProGlyThrMetGly 50
55014 GGGCGCGGACCGCACTTGCCCTCCATTTGCCCGCGGAGTAAGGCG 55063
51 PheProPhePheGlyGluThrLeuGlnMetValLeu.Gln..... 63
55064 TTCCCTCTTTGGGGAAACCTTGACAGATGCTCAGGTAGGAGGAGG 55113
63 ..... 63
55114 TGGGGGGGACAGCGCTGCTCCCGGAGCGCGGCGCTGGCTTCT 55163
63 ..... 63
55164 GCTGAAGTCGGGGTAGGCGCCCGGAGGAGCATGTATTGGCGTAGAG 55213
63 ..... 63
55214 CAGGGCTGGCGGAGCGGCGCTCCCGGCGCCCTCATGGCCACTTC 55263
64 ..... 64
55264 TCTCTCCGCTTCTCCACAGCGGAGGAGATCTCTGCAGATGAAGCC 55313
73 ArgGlySTyrGlyPheIleTyrIleThrHisLeuPheGlyArgProThyA 89
55314 AGGAATACGGGCTTCATCAAGAGCGATCTGTCGGCGGCCACCGT 55363
89 LarGylMetGlyAlaAspAsnValArgArgIleLeuLeuGlyAspAspA 106
55364 ACGGATGATGGCGCGGACAAATGTGGGCGCATCTTGCTGGAGACACC 55413
106 rGleuValSerValHisTrpProAlaSerValArgThrIleLeuGlySer 122
55414 GGCTGATGTCGGTCCACACAGCGGCGGTGGGACCATTTCTGGGATCT 55463
123 GlyCysLeuSerAsnLeuHisAspSerHisIleGlnArgLys.Lys. 138
55464 GGCTGCTCTCTTAACCTGCACGACTCTCGACAAAGCAGCAGAAAGGT 55513
```

```
138 ..... 138
55514 GGGGGGACAGAGCGGACGCTGACAGGAGGAGGAGCCCATTTATAGCG 55563
138 ..... 138
55564 GAATTCGGGTGATGATGCTAGCGCGGGCTAGCACTTGAGTGGGCT 55613
138 ..... 138
55614 AGGACCTCTGCCAGCTCCAGTTAGCTATACCACTCGAGAGTGCAC 55663
138 ..... 138
55664 TGTGTGGCAGACTGGGGGTCTTGAAAGGGAGCGCGGTAGACGACA 55713
139 ..... 139
55714 GGGCGGATGAGGCTTTTAACGCTGTCCCTCTCGGAGCTCAGGTGAT 55763
140 eMetArgAlaPheSerArgGluAlaLeuGlyCys.....ValI 140
55764 TATCGGGCTTCAGCGCGGAGGACTCGAATGCTACGTGCCGTGATCA 55813
151 ..... 151
55814 CCGANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 55863
151 ..... 151
55864 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 55913
152 ....TyrValProValIleThrGluGluValGlySerSerLeuGluNT 167
55914 NNNNTACGTCGCCGCTGATCACCGAGGAAGTGGGACAGCTGAGACAG 55963
167 rPLeuSerCysGlyGluArgGlyLeuLeuValTyrProGluValLysArg 183
55964 GGCTGAGCTGGCGGAGCGCGGCTCTGCTGCTACCCCGGAGGAGCGG 56013
184 LeuMetPheArgIleAlaMetArgIleLeuLeuGlyCysGluProGlnE 200
56014 CTGATTTCCGATTCGCGATGCGCATGCTACTGCTGCTGCGAACCACACT 56063
200 uAlaGlyAspGlyAspSerGluGlnGlnLeuValGluAlaPheGluGluM 217
56064 GCGGCGGAGGAGGAGCTCCGAGCAGCAGCTTGTGGAGGCTTGAGAGAA 56113
217 eThrTrpAsnLeuPheSerLeuProIleAspValProPheSerGlyLeu 233
56114 TGACCCGCAATCTTCTGCTGCTGCCATCGACGTGCCCTTCAAGGGGCTG 56163
234 TyrArg..... 235
56164 TACCGGCTAAGGCGGCAAAAGGCTGCGGACTAGGCGCGGAGCTGG 56213
236 ..... 236
56214 GGCTGCTCACCGCGCGCGCTCTGCTGCGCTCAGCGGATGAAGGCGCG 56263
241 AsnLeuIleHisAlaArgIleGluGlnAsnIleArgAlaLysIleCysG 257
56264 AACCTTATTCAGCGCGGCTCGAGCAGAAACATTTGCGGCAAGATCTG 56313
257 yLeuArgAlaSerGluAlaGlyGlnGlyCysLysAspAlaLeuGlnLeu 274
56314 GCTCGCGGATCGAGAGGCGGCGAGGCTGCAGAAAGCGGCTGACACTCT 56363
274 euIleGluHisSerTrpGluArgGlyGluArgLeuAspMetGln..... 288
56364 TGATGAGCACTCTGCGGAGAGGAGGAGCGGCTGACATGACGTGAGT 56413
288 ..... 288
```

56414 AGCAGCTTCAGACCAAGCACTGGGAGTTTGGTCCCTGGCTTCCAAAGC 56463
 288 288
 56464 GCTGTCTGGGGCCCCCAAGCGCGCTGGGGCCCAAGCTTCTTGAG 56513
 288 288
 56514 TGGCGCGCGGCTCAGACTACAGCTATGGAATCCGAGGAAGGTGAGA 56563
 288 288
 56564 CACCGGTACAGAGAGCTGGGAAGGGCTGGCGCGAACTGGAGCAATC 56613
 288 288
 56614 CCTAGCCTTCCAGTTTCAAGGAAAGTGGAAATTGCAAAAATGTT 56663
 288 288
 56664 AATAAGAACCTTGCATTTTAAATAAATAAGACTTAAGCTTACAGAGTT 56713
 288 288
 56714 TCCGTTAGAGCGGGGTCTACTGCTTACTGCTCCAGCTGAATAAGG 56763
 288 288
 56764 GAGCTGCATTTTGTAAAGATATGCTTTCCTTGACTTCTGTGACGA 56813
 288 288
 56814 AAACATTAGCCCTTCTAGTCTTCCCTCCAGAACTCTAGTGCATTCTG 56863
 288 288
 56864 AGTATGCTTCTGCAACCGCAGGACAGCTTGAGAAATGTGGTCTCA 56913
 289 Ala..LeuLysGlnSerSerThrGluLeuLeuPheGly 301
 56914 CTCATTTCTTGAAGCAATAAGCAATCTTCAACGAACTCCTCTTGAGG 56963
 301 YHISGluThrThraIAserAlaIathrSerLeuIleThrTyrLeuGly 318
 56964 ACAGGAAACCAAGCGCAGTGCAGCCACATCTGTATCATCTTACCTGGGC 57013
 318 euTYRProHisValLeuGlnLysValArgGluGluLeuLysSerLys... 333
 57014 TCTACCCACATGTTCTCCAGAAAGTGCAGAAAGCTGAAGATTAAGSTA 57063
 333 333
 57064 GGGAGACTGGGCTGGGGGTGCTTATTAGCTTAGCAAAATTCAGTGC 57113
 333 333
 57114 CCTTAGCAACTTCGAATAAGTCAAGTGTGCTGCCTTCATGAGATATTT 57163
 333 333
 57164 GAAAGTGAAGCCAGGCTGCCGTGGGCCAGTGGGCAAGAAATAGCTTG 57213
 333 333
 57214 TGAATAAACAGAGATGAGTCTTGAAGTGTGAGCCCACTTGGGCAAGGTTT 57263
 333 333
 57264 GACCTTGTATTCCTAACCCCTCTCCACCTTTTGTGAACCGTGGAGATTC 57313
 333 333

57314 TCAGATAGTTCCACTTCTTGAATGGTGTGTCCAAAGGCATACATAGT 57363
 333 333
 57364 GTGAATAGCTGATTAGTGTGGGTGGTGTGATGTGGGGTGGGGTT 57413
 333 333
 57414 TTTTGGGGGAACCTTGTCCTTCTTACTGTGAGCAGCATCTCTGGGA 57463
 333 333
 57464 TTGTAATATGATAGTGAATTTGGCAGGCAATGGCATTAGCTGTGT 57513
 333 333
 57514 TCCCTCCAGCTCTCCCATCATGAGGCCCTTGGGTTAGTCTCAC 57563
 333 333
 57564 TTAAGCCCTGTAGCTCTGCTGGGCTGATTTATTGGAGCAAAATTA 57613
 334 Gly..LeuLeuCysLysSerAs 340
 57614 CTGTTACCTCTGTATGACTGTTTGATAGGCTTACTTTCGAAAGAGA 57663
 340 nGlnAspAnLysLeuAspMetGluIleLeuGluLeuLysTyrIleG 357
 57664 TCMAACAACAAAGTTGACATGGAATTTTGGAAACACTTAATATCATCG 57713
 357 LysValIleLysGluThrLeuArgLeuAsnProProValProGlyGly 373
 57714 GGTGTATTATTAAGAGACCTTCGACTGAATCCCCAGCTCCAGAGGG 57763
 374 PheArgValAlaLeuLysThrPheGluLeuAsn... 384
 57764 TTYCGGGTTCCTCTGAAGACTTTTGAATTAAGTAAATTTCTCT 57813
 384 384
 57814 TCTTTCTCCTTTGTTGGTGAATAAACTCTCTTCTTCCATGCT 57863
 384 384
 57864 GTGGCTGCATTTCTTATGCTTTGATAATGTTCTGCCATATGAGAT 57913
 384 384
 57914 ATGTTTCAGCAACCTGATCCACCTCTCTTCTACTCTCCCTTGCCT 57963
 384 384
 57964 TTAGCTCATATTTCCCCCAAGATATCAGTGCCTTTTGTGTGA 58013
 385 Gly 385
 58014 AAGTTAATTCAGTGTGTCAGCCCTGGGTTTCATATCTTTTGCAGG 58063
 386 ..TyrGlnIleProLysGlyTyrAsnValIleTyrSerIleCysAspThr 401
 58064 GATACCAAGATTTCCCAAGGCTGGAATGTATCTACATATCTGTGATCT 58113
 402 HisAspValAlaGluIlePheThrAsnLysGluGluPheAsnProAsp 418
 58114 CATGATGTGGCAGAGATCTTACCAACAAGAAATTTATCTGTGACG 58163
 418 gPheSerAlaProHisProGluAspAlaSerArgPheSerPheIlePro 435
 58164 ATTCACTGCTGCCTCACCCAGAGATGCATCCAGGTTCAAGCTTCAATTCAT 58213
 435 heGlyGlyGlyLeuArgSerCysValGlyLysGluPheAlaLysIleLeu 451
 58214 TTTGAGGAGGCTTAGGAGCTGTGTAGGCAAGAAATTTGCAAAAATTTCTT 58263

452 LeuYsIlePheThrValGluLeuAlaArgHisCysAspTrpClnLeu 468
 |||||
 58264 CTCAAATATTTTACACTGAGCTGGCCAGGCATGTGACTGGCAGCTTCT 58313
 468 uAnsglyProProthrMetLysThrSerProthrValTyrProValAspa 485
 |||||
 58314 AATGAGACCTCTTACATGAAACAGTCCACCGGTATCTCTGTGGACA 58363
 485 snLeuProAlaPheThrHisPheHisGlyGluIle 497
 |||||
 58364 ATCTCCCTGCAAGATTCCACCATTTCCATGAGGGAATC 58401
 seq_name: gb_cov:AF199462
 seq_documentation_block: 1479 bp mRNA VRT 21-NOV-1999
 LOCUS AF199462
 DEFINITION Gallus gallus retinoic acid degrading enzyme CYP26 mRNA, complete
 cds.
 ACCESSION AF199462 GI:6456740
 VERSION AF199462.1
 KEYWORDS chicken.
 SOURCE Gallus gallus
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 Phasianinae; Gallus.
 REFERENCE 1 (bases 1 to 1479)
 Swindell,E.C., Thaller,C., Sockanathan,S., Petkovich,M.,
 Jessell,T.M. and Eichele,G. Complementary Domains of Retinoic Acid Production and Degradation
 In the Early Chick Embryo
 Dev. Biol. (2000) in press
 2 (bases 1 to 1479)
 Swindell,E.C., Thaller,C., Sockanathan,S., Petkovich,M.,
 Jessell,T.M. and Eichele,G. Direct Submission
 Submitted (28-OCT-1999) Max Planck Institute for Experimental
 Endocrinology, Feodor-Lynen Strasse 7, Hannover 30625, Germany
 TITLE JOURNAL
 AUTHORS Swindell,E.C., Thaller,C., Sockanathan,S., Petkovich,M.,
 Jessell,T.M. and Eichele,G.
 FEATURES
 source 1. 1479
 /organism="Gallus gallus"
 /db_xref="taxon:9031"
 1. 1479
 /note="p450RA1"
 /codon_start=1
 /product="retinoic acid degrading enzyme CYP26"
 /protein_id="AA09250.1"
 /db_xref="GI:6456741"
 /translation="MGPSALVASALCTFLPLILFLAAYRLMDLYCAGSRDPSCLPL
 PPTGMLPRGERTLQWYLRKFLQMKRRKTKYTHLGRFVRVGMENVRHLL
 GERRLYSVQMPASVKTILGSLNHNHGHKRRKVTMOAFSDALQHVAVYQEEV
 SACLQWLGAPCLLYPEVKRLMFRLMILGFDRQASPGEDOLVAFEBMIN
 LEPLIDVPSGTYRGLRARNIITIAKEENIRAKMAKEEGGKVKRELQIMETOGN
 GEOLNQELKESATELLFGHETLASATSLIAFLGHHVDLQKRELQKGLSP
 NQKQALMEFLQELKTYGCVIKETLRLSPVPGFRALATLELNQIDKGNVVIS
 ICDTHVADLFMDKDEPNRPMSPSPEDSSRSFTFGGSLSCVSKERAKVLAKIE
 TVLASLSCMDQLNCPPTMKTRIVYEDVNLPAKFIQSGQI"
 BASE COUNT 302 a 438 c 427 g 312 t
 ORIGIN
 alignment_scores:
 Quality: 2030.50 Length: 497
 Ratio: 4.463 Gaps: 3
 Percent Similarity: 91.549 Percent Identity: 78.873
 alignment_block:
 US-09-668-482-4 x AF199462 ..
 Align seg 1/1 to: AF199462 from: 1 to: 1479
 1 MetGlyLeuProAlaLeuAlaSerAlaLeuGlyThrPheValLeuPr 17

|||||
 1 ATGGGGCTTCGCGCTGTGTCGCGACGGCCGTGACACCTTCCTGCTGCC 50
 17 oLeuLeuLeuPheLeuAlaAlaIleLysLeuThrPaspLeuTyrCysValS 34
 |||||
 51 CCGCGCTCTTCTCTGCGCGCGCTGCGGCTGCGACCTGTACTGTGGCCA 100
 34 eGlyYArgAspArgSerCysAlaLeuProLeuProGlyThrMetGly 50
 |||||
 101 GCGCGCCGACCCAGCCTGCTCCGCTCCGTCGCCCGGCGACCATGGGG 150
 51 pheProPhePheGlyGlyuThrLeuGlnMetValLeuGlnArgArgLysPh 67
 |||||
 151 CTGCGCTTCTGCGGAGACGCTGAGATGTGCTGACGCGCGGGAATT 200
 67 eLeuGlnMetLysArgArgLysTyrGlyPheIleTyrLysThrHisLeuP 84
 |||||
 201 CCGGAGATGAAGCGGAGAAATACGGCTTCATCTACAAGACTCACCTCT 250
 84 heGlyArgProThrValArgValMetGlyAlaAspAsnValArgArgIle 100
 |||||
 251 TTGGGGGCGCCACTGTGCGGTGATGGCGCTGAGAAATGTGCGGCATC 300
 101 LeuLeuGlyAspAspArgLeuValSerValHisTrpProAlaSerValAr 117
 |||||
 301 CTGCTGGGCGACACCGCGCTGCTGCTGAGTGGCCGCGCTCGGTGCG 350
 117 gThrIleLeuGlySerGlyCysLeuSerAsnLeuHisAspSerHisL 134
 |||||
 351 CACCATCTTGCGGTGGGCTGCTCTCCAACTTCACACGGGCGACACCA 400
 134 yGlnArgLysLysValIleMetArgAlaPheSerArgLysAlaLeuGln 150
 |||||
 401 AACACCGCAAAAGGATGATGACGACCTTCCCGGAGATGCCCTGCGAG 450
 151 CysTyrValProValIleThrGluGluValGlySerSerLeuGlnItr 167
 |||||
 451 CACTAGTGCTGCTCATCCAAAGAGGTGAGTGCCTGCTGGACAGAGT 500
 167 pleuSerCysGlyGlyuArgGlyLeuLeuValTyrProGluValLysArgL 184
 |||||
 501 GGTGGGTGCGGGGCC...TGCTGCTGCTGATATCCGAGGTGAAGCGCC 547
 184 eumetPheArgIleAlaMetArgIleLeuLeuGlyCysGluProGlnLeu 200
 |||||
 548 TCATGTTCGCGCATGCGCATGCGCATGCTGCTGCTGCGCCCGCCAG 597
 201 AlaGlyAspGlyAspserGluGlnGlnLeuValGluAlaPheGluGln 217
 |||||
 598 GCGAGCCCC...GATGCTAGCAGCAGCTGGTGGAGCCTTGAGAGAGAT 644
 217 ThrArgAsnLeuPheSerLeuProIleAspValProPheSerGlyLeu 234
 |||||
 645 GATCGCAACCTCTTCTCCCTCCCATGATGCTGCTTCACTGGGCTCT 694
 234 YArgGlyMetLysAlaArgAsnLeuIleHisAlaArgIleGluGlnAsn 250
 |||||
 695 ACCGGGGCCCTGAGGGCGGCAACATCATCCATCCACAGATGAGGAAAC 744
 251 IleArgAlaLysIleCysGlyLeuArgAlaSerGluAlaGlyGlnIcy 267
 |||||
 745 ATCGGTGCCAAG.....ATGGCCCGCAAGAGCCCGAGAGGTGGCTA 785
 267 sLysAspAlaLeuGlnLeuLeuIleGluHisSerTrpGluArgGlyGluA 284
 |||||
 786 CAAGGATGACATGACGCTGTTGATGAGACACACAGGATATGAGAGAC 835
 284 rGluAspMetGlnAlaLeuGlyGlnSerSerThrGluLeuLeuPheGly 300
 |||||
 836 ACCTCAACATGCGAGAGCTGAAGAGTCTGCCACAGAGTCTCTTTGGA 885
 301 GlyHisGluThrThrAlaSerAlaAlaThrSerLeuIleThrTyrLeuG 317
 |||||

	886	GGTCAACAAACACAGGCCAGTCTGCTCATCATTACTGATTCGCTTCCTAGG	935
	317	yLeuTyPrOHnIsValIeugInLyVaLaArgGuLubLeuLySerLySg	334
	:::	:::	:::
	936	GCTCCACACATGATGTTCTGCAGAAAGTGAGGAAGAGCGTCAATTAAGA	985
	334	IyLeuLeuCysLySSerAngInAsPaSnLySleuAsPMeGLuILEu	350
		:::	:::
	986	GGTACTGCTGCTGCTCCACCAAGAAAAACAATTBAACTAGAGCTCTTG	1035
	351	GUUGuLeuLyStyrILeGIcYsValILeLySGluTrPheLeuArgLeuAs	367
			:::
	1036	GACACACTGAAGTACACAGCGCTGTGTATCACAGAAAACCCTCAGGCTAG	1085
	367	nProProValProGIyGIyPheaArgVaLaLeuLySThrPheGluEua	384
		:::	:::
	1086	CCTCCGCTTCCTGAGAGATTCGGAATTCCACTCAAGACCCCTTGAGCTGA	1135
	384	sndLytyrGlnILePolysGIyTPRAsnValILeTySerILecYasP	400
	1136	ATGGTTAACAGATCCCCAAGGCTGGAGATGTTATTATTCACGATCTGCAT	1185
	401	ThriHisArvAlaIagLuILePheThrAnLySGluGluPheAsnProAs	417
		:::	:::
	1186	ACCACAGATGTGGCCGATCTTTACCGACAAGACGATTCAAACCCGGA	1235
	417	pArgPheSerAlaProHisProGluSprLaSerArgPheSerPheIleP	434
		:::	:::
	1236	CCGGTTATATGTCTCCGCTCTCCGGAGAGCTCTCTGAGTTCAAGTTTCATTC	1285
	434	roPheGIyGIyGIyLeuArgSerCySvaIGLYSGluPheALaLySile	450
		:::	:::
	1286	CTTTCGGTGGGGGCTTGAGAGCTGCGTGAAGCAAAGATTTGCCAAAGTCC	1335
	451	LeuLeuLySLePheThrVaLIguLeuAlaArgHISCyAsPTripSile	467
	1336	CTTGTAAAAATTATTACGTGAGCTGGGCGCGGAGCTGTGACTGGGAGCT	1385
	467	uLeuAsnGIyProProThrMetLySThrSerProThraLyTyProVala	484
		:::	:::
	1386	GCTAAATGGACCTCTCATCATGAAGACGGGCCCATAGTACCTGCTGG	1435
	484	sPaSnLeuProAlaArgPheThrHisPheHisGLyLUtle	497
	1436	ACAATCTGCTCCGCCCAATTCATAGGTTTCAGCGGCCCAATTC	1476
seq_name:	gb_ov:	AF057566	
seq_documentation_block:	1479 bp	mRNA	VRT
LOCUS	AF057566		10-MAR-1999
DEFINITION	Xenopus laevis retinoic acid converting enzyme mRNA, complete cds.		
ACCESSION	AF057566		
VERSION	AF057566.1 GI:3283171		
KEYWORDS	.		
SOURCE	African clawed frog.		
ORGANISM	Xenopus laevis		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae;		
	Xenopodinae; Xenopus.		
REFERENCE	1 (bases 1 to 1479)		
AUTHORS	Holleman,T., Chen,Y., Grunz,H. and Pieler,T.		
TITLE	Regionalized metabolic activity establishes boundaries of retinoloic acid signalling		
JOURNAL	EMBO J. 17 (24), 7361-7372 (1998)		
MEDLINE	99077803		
REFERENCE	2 (bases 1 to 1479)		
AUTHORS	Chen,J.L., Grunz,H., Panltz,F., Pieler,T. and Hollemann,T.		
TITLE	Direct Submission		
JOURNALT	Submitted (05-PR-1998) Developmental Biochemistry, University of Goettingen, Humboldtallee 23, Goettingen, Lower Saxony 37073, Germany		
FEATURES	Location/Qualifiers		
source	1..1479		

[illegible]

501	GCTCAGAGCGGGCCCC...TGTGTCGTGGGTATCCCGCATCAAGAA	547
184	eumetpnehtglllealmetargllleuleuclgysgluproglnleu	200
548	TGATGTCCTGGATGGCATAGAGCTCTGCTGGCTGTATCCCAACG	597
201	AlaGlyAspGlyAspSerGlnGlnLeuValGlnAlaPheGlnGlu	217
598	ATG...GACGGCAACAGAGGAACGCTCTGMAACCTTCAGAGAAAT	644
217	tThrArgAsnLeuPheSerLeuProIleasPvalProPheSerGlyLeuT	234
645	GAGCGGAATCTTTCTTCTTGGCTATGTATGTGCGGTTAAGCGGCTCT	694
234	YFAFGYMeLysAlaArgAsnLeuIleHisAlaArgIleGlnIleAsn	250
695	ACGGGGCTGTGGGGCCAGAAACCTTAATTCATCCCAATTTGAGAAAC	744
251	lleargAlaLysIleCysGlyLeuArgAlaSerGlnAlaGlyGlnGlyC	267
745	ATCAAGAAAGAAGCTG.....CAAGAGAACAGATGAACATTG	782
267	slyAspAlaIleuGlnleuIleuIleGlnHisSerTrpGlnArgGlyGlnA	284
783	CAAGGATGCCCTGAGTGTGATTAATAATACCCGACAGAAATGGCGAC	832
284	rgLeuAspMetGlnAlaLeuLysGlnSerSerThrGlnleuIleuPheGly	300
833	CAATCAACTTGCAGAGCATTAAGAAAGTCCGCAACATGAGCTCTGTTGGA	882
301	GlyHisGlnThrThrAlaSerAlaAlaThrSerLeuIleThrTyrLeuG	317
883	GGGCACGAAACCAAGCAAGTGTGCGACGTCCTCACTCTTTCTTCTGC	932
317	YLeuTyrProHisValIleuGlnLysValArgGlnIleuLysSerLysG	334
933	GCTTCACAAAGACGTTCTGGAAAGGTCGCAAGAGGCTTGAAACACAG	982
334	lyLeuLeu...CysLysSerAsnGlnAspAsnIleuAspMetGlnIle	349
983	GTCCTGTGTCACATAAACCCGAGAAAGAAAGAACTGAGCATCGAATGA	1033
350	LeuGlnIleuLysTyrIleGlyCysValIleLysGlnThrLeuArgLe	366
1033	CTTCAACAGTTAAAGTAATCTGACTGTGCATTAAAGAAACCTCGGCT	1083
366	uasProProValProGlnGlyPheArgValAlaLeuLysThrPheGlnL	383
1083	CACCCACACAGTGGCGGATGATTCGGCTGCTCAAGACGTTTGAT	1133
383	euaGnGlyTyrGlnIleProLysGlyTyrAsnValIleTyrSerIleCys	399
1133	TAAATGGCTATCGATTTCCCAAGGGCTGAATGTAATATACAGCATTC	1183
400	AspThrHisAspValAlaGlnIlePheThrAsnLysGlnIlePheAsnPr	416
1183	GACACCCACGAGTGGCGCATGTGTTCCGACACGAGCAACAATTCAACC	1233
416	asPArgPheSerAlaProHisProGlnAspAlaSerArgPheSerPheT	433
1233	TGACCCCTTTTAAACCCACTCCAGAAATCTCCCAAGGTTTGCTGTCA	1283
433	leProheGlyGlyLeuArgSerCysValGlyLysGlnPheAlaLys	449
1283	TCCCTTCGAGAGTGAAGTGAAGTGTGCTGATGGCAAGAAATTTCTTAA	1333
450	lleLeuLeuLysIlePheThrValGlnLeuAlaArgHisCysAspTrpG	466
1333	ATCCTTCTCAAGGTGTGTGGTGAATGTGTCTAATTGGAGACTGGGA	1383
466	lleuLeuAsnGlyProProThrMetLysThrSerProThrValTyrProV	483
1383	GCTCTGAAACGGATCCCGCCATACAACTCCCGATTTATCTGTGCTG	1432

```

483 alspanleuproalaarGphenhrisPheHISGLYGVIIIE 497
|||||.....|
1433 TGGACACCTACCCGCTTAATTCACAAACCTTTCACACTTATT 1476

seq_name: gb_ov:DR068234

seq_documentation_block:
LOCUS DR068234 1479 bp mRNA VRT 22-NOV-1996
DEFINITION Danio rerio all-trans-retinoic acid 4-hydroxylase (p450RAI) mRNA,
ACCESSION U68234
VERSION U68234.1 GI:1680715
KEYWORDS
SOURCE zebrafish.
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Rasbora; Danio.
REFERENCE
1 (bases 1 to 1479)
White,J.A., Guo,Y.D., Baetz,K., Beckett-Jones,B., Bonasoro,J.,
Hsu,K.E., Dillworth,F.J., Jones,G. and Petkovich,M.
Identification of the retinoic acid-inducible all-trans-retinoic
acid 4-hydroxylase
J Biol. Chem. 271 (47), 29922-29927 (1996)
J97094702
2 (bases 1 to 1479)
White,J.A., Guo,Y., Baetz,K., Beckett-Jones,B., Bonasoro,J., Hsu,K
E., Dillworth,F.J., Jones,G. and Petkovich,M.
Direct Submission
Submitted (28-AUG-1996) Cancer Research Labs, Queen's University,
Botterell Hall, Kingston, ON K7L 3N6, Canada
FEATURES
Source
1..1479
/organism="Danio rerio"
/db_xref="taxon:7955"
1..1479
/gene="p450RAI"
1..1479
/gene="p450RAI"
/note="all-trans-retinoic acid 4-hydroxylase"
/codon_start=1
/protein_id="AACG0045.1"
/db_xref="GI:1680716"
/translation="MGLTLMVFLCTIVLPVLEIAVAKLWEMIMIRVDPNCRSPIL
PGRMGLPGLTIGTILIORRKLRMKROKGCYKHHNENPVRKMGADNYOILIL
GEHLVAYOMPASVRIILGSDTSLNVEGVOKNNKKIMRAFSDALEHTIPVIOEE
KSAIQEWLQKDSVLYIPEMKLMFRIAMRIILGLFEBQIKTDQELVEAEEMIKIN
FSLPIDVPSGLYGLRANRINRSKIEENIRKLQDDNENRKYKALQILNRSR
SDPEFSIQAMKEAATELLFGHEHTTASTASLWFLINTEVGVREVEQEKEMGK
YTPCKGLSMELLDQIKYGCYIKETLRINPVPGLFALKTPELNGQIPKGNVYI
SICHTGVADVPEKKEFQEPERFMSKGLDGSERNYIPFGGSRMCGKBEFAKVLKIK
FLVELDHCNMWILISNGPPTKKTGTPTVPVDMLPKFTSYAN"
BASE COUNT 425 a 331 c 379 g 344 t
ORIGIN
alignment_scores:
Quality: 1738.00 Length: 497
Ratio: 4.080 Gaps: 5
Percent Similarity: 85.714 Percent Identity: 68.008
alignment_block:
US-09-668-482-4 x DR068234 ..
Align seg 1/1 to: DR068234 from: 1 to: 1479
1 MetGlyLeuProAlaLeuLeuAlaSerAlaLeuGlyThrPheValLeuPr 17
|||||.....|
1 ATGGGGCTGTACACCTTATGGTCACTTCTCTGTGACCATGTGCTAC 50
17 OleuLeuLeuPheLeuAlaAlaIleLysLeuTrpAspLeuTYrCysValS 34
|||||.....|

```



```

/codon_start=1
/product="cytochrome P450 retinoid metabolizing protein
P450RAI-2"
/protein_id="AAF76003.1"
/db_xref="gi:8515441"
/translation="MLEGLDVLVSALATLACLVSTVLLAVSQOLWQLRMAATRDKS
CKLPIPKSMGFPLIGETGHWLLOGSGFSSRRREYGVFTHLIGRPLIRVTGAEVY
RKILMGHHLVSTEMRSTRLMIGPNTVSNIGDIHNRKRYFSKIFSHALSYLPK
IQLVIOPLRAMSSHPKAIINYOEAKLTFPMALIKFVLGFSIPEDLGHLEVYQOFV
DNVSLPYDLPEFSGVRGILQROILQGLKRAIRKILQCTGQKDYDLADLLITSSKE
HGKEMTQELKDGTELLELFAVATTAASSTLIMOLKHPVLEKRLRELAHGLIHS
GGCECEGLRDLTSLGLRLDCVLRKVRLETPISGGRTVLYQTELDGFQPKGWSV
MSTRDHDIAFVRKDVNPDRPSQARSEDKGRFHYVLPFGGVRICLKHAKLF
LKVLAVELASTREFLARTFPRITLVPLVHPVGLSVKFFGLDSMONEILPETEAML
SATV"
BASE COUNT      893 a      1265 c      1240 g      1047 t
ORIGIN

alignment_scores:
    Quality: 1041.50      Length: 495
    Ratio: 2.838      Gaps: 9
    Percent Similarity: 74.141      Percent Identity: 44.242

alignment_block:
US-09-668-482-4 x AAF252297 ..
Align seg 1/1 to: AAF252297 from: 1 to: 4445

7 LeuAlaSerAlaLeuGlyThrPheValIleuProLeu.....Le 19
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
22 CTGGTGTGGCGGTGGCCACCTCGCGCGGTGCTGTGCTGCTGCGTGC 71
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
19 uLeuPheLeuAlaAlaIleLysLeuTrpAspLeuTrpCysValSerGly 36
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
72 GCTGGTGGCGGTGCTGCGACAGCTGTCGACGCTGCGCTGGCGCCAC 121
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
36 rGAspArgSerCysAlaLeuProLeuProGlyThrMetGlyPhePro 52
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
122 GCGACAGAGCTGCACAGCTGCACCTCCCAAGGATCCATGGGCTTCCG 171
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
53 PhePheGlyGluThrLeuGlnMetValIleuGlnArgArgLysPheLeu 69
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
172 CTCATCGGAGAGACCGGCCACCTGGCTGCTGCAGGGTTCGCTTCCAG 221
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
69 nMetLysArgArgLysTyrGlyPheIleTyrLysThrHisLeuPheGly 86
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
222 GTCGGGAGGAGGAAGTATGGCAACGCTTTCAAAGCAGCATTTGTTGG 271
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
86 rGProThrValArgValMetGlyAlaAspAsnValArgArgIleLeuLeu 102
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
272 GGCCCTGATACGCGTGCACGGCGGAGAAACGTCGGCAAGATCCTCAG 321
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
103 GlyAspAspArgLeuValSerValHisTrpProAlaSerValArgThrI 119
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
322 GCGGAGCACACACTCGTGAGCAGCAGATGGCCTCGACAGACCCGATGT 371
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
119 eLeuGlySerGlyCysLeuSerAsnLeuHisAspSerSerHisLysGln 136
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
372 GCTGGGCCCCCAACAGCGTGTCCATTGGCGACATCCACCCACACA 421
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
136 rGlyLysValIleMetArgAlaPheSerArgGlnAlaLeuGlnLysTyr 152
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
422 AGCCGAAGCTTCTTCCAAGATCTTCAGCCACAGGCGCTGGAGATTAC 471
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
153 ValProValIleThrGluGluValGlySerSerLeuGlnTrpLeuSer 169
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
472 CTGCCAAGATCCAGCTGTGATCCAGACACACTGGCGCTGGAGACAG 521
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
169 rCysGlyLysLysGlyLeuLeuValTyrProGluValLysArgLeuMetP 186
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
522 CCACCCCGAG...GCCATCAACAGTGTACAGAGGCGCAGAAAGTGA 568

```

```

186 heArgIleAlaMetArgIleLeuLeuGlyCysGlu...ProGlnLeuAla 201
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
569 TCCGCATGGCCCATCCGGGTCTGCTGGGCTTCACAGATCCCTGAG... 612
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
202 GlyAspGlyAspSerGluGlnIleValAlaAlaPheGluGluMetTrp 218
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
613 .....GAGACCTTGGGACACCTCTTGAAGTGTACACAGAGATTGT 653
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
218 rArgAsnLeuPheSerLeuProIleAspValIleProPheSerGlyLeuTyr 235
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
654 GAGACATGTTCTTCTCCCTGCTGTCGACGCTGCTTCAGAGGCTACCG 703
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
235 rGlyMetLysAlaArgAsnLeuIleHisAlaArgIleGluGlnAsnIle 251
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
704 GGGGCAATTCAGGCTCGGCANATCTCGAAGAGGGCTGGAGAGGCCATC 753
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
252 ArgAlaLysIleCysGlyLeuArgAlaSerGluAlaGlyGlnGlyCys 268
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
754 CCGGAGAGAGCTG.....CAGTGCACACAGGCGCAGAGACTACTT 791
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
268 sAspAlaLeuGlnLeuLeuIleGluHisSerTrpGluArgGlyGluArg 285
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
792 GAGACCCCTGACCTCTCTCATTTGAGACACAGACAGACGAGAGAGAGA 841
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
285 uAspMetGlnAlaLeuLysGlnSerSerThrGluLeuLeuPheGlyGly 301
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
842 TGACCATGACGAGCTGAAGAGAGGAGACCTGAGCTGATCTTTGGCG 891
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
302 HisGluThrThrAlaSerAlaAlaThrSerLeuIleThrTyrLeuGly 318
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
892 rATGCCACACAGGCGACGCGCACCTCTCATCATGATGACGCTGTGAA 941
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
318 uTyrProHisValIleuGlnLysValArgGluGluLeuLysSerLysGly 335
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
942 GCACCCCACTGTGCTGAGAGAGCTGGGGATGAGCTGGCGGCTCATG 991
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
335 eLeuLeu.....CysLysSerAsnGlnAspAsnLysLeuAspMet 347
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
992 TCCTGACAGATGGCGGCTGCCCTGC.....GAGGCGACACTGCGCTG 1035
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
348 GluIleLeuGlnIleuLysTyrIleGlyCysValIleLysGluThrLe 364
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1036 GACACGCTAGTGGCTGCGCTACCTGAGCTGCGTATCAAGAGAGTCA 1085
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
364 uArgLeuAsnProProValProGlyGlyPheArgValAlaLeuLysThr 381
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1086 GCGCCTGTTACAGCCCATTTCCGGGGCTTACCGCACTGTCGACAGCT 1135
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
381 heGluLeuAsnGlyTyrGlnIleProLysGlyTyrAsnValIleTyrSer 397
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1136 TCGAGCTGTATGTTCCAGATCCCAAGAGCTGAGTGTATGATAC 1185
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
398 IleCysAspThrHisAspValAlaGluIlePheThrAsnLysGluGlu 414
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1186 ATCCGGGACACCATGACACAGCGCCGCTGTTCAAGACGTAAACCTGT 1235
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
414 eAsnProAspArgPheSerAlaProHisProGluAspAla...SerArg 430
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1236 CGACCCCGATTCCTTACAGCAGCGGAGAGAGAGCAAGATGGCGCT 1285
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
430 heSerPheIleProPheGlyGlyLeuArgSerCysValIleLysGlu 446
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1286 TCCATTACCTCCCTTCCGTGGCGGCTCCGAGACCTGCTGGGCAAGC 1335
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
447 PheAlaLysIleLeuLeuLysIlePheThrValIleGluAlaArgHis 463
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1336 CTGGCCAACTTCTTCTGAGGTGCTGGCGGAGAGCTGCTGACACCA 1385
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
463 sAspTrpGlnLeuAsnGlyPro...ProThrMetLysThrSerProT 479
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1386 CCGCTTGTAGCTGCTACACGAGACCTTCCCGGCAATCACTGTGTC 1435
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
479 hValTyrProValAspAsnLeuProAlaIleArgPhe 490

```


1436 TCCTGCACCCCGGATGCGCTCAGCGTCAAGTTC 1470

seq_name: gb_ba3:SYCSLLE

seq_documentation_block:
LOCUS SYCSLLE 113064 bp DNA BCT 13-FEB-1999
DEFINITION *Synechocystis* sp. PCC6803 complete genome, 22/27, 2755703-2868766.
ACCESSION D64003 AB001339
VERSION D64003.1 GI:1001200
KEYWORDS tRNA-fMet; 2-keoacetyl dehydrogenase (malate dehydrogenase, lactate dehydrogenase); 30S ribosomal protein S18; 50S ribosomal protein L33; 6-aminohexanoate-cyclic-dimer hydrolase; Kdtb; Mg-protoporhyrin IX monomethyl ester oxidative cyclase 66 kD subunit; MoxK protein; NADH dehydrogenase subunit 5; MoxK; Plad; acetyl-coenzyme A synthetase; acyl-CoA desaturase 1; alanine racemase; aspartate 1-decarboxylase; cyanate lyase; cytochrome P450; dtdp-glucose 4,6-dehydratase; elongation factor EF-G; endo-1,4-beta-glucanase; ferredoxin-nitrite reductase; ferredoxin-thioredoxin reductase; catalytic chain; ferredoxin; flavoprotein; low affinity sulfate transporter; methionine aminopeptidase; methyltransferase; molybdenum cofactor biosynthesis protein A; molybdenum cofactor biosynthesis protein C; molybdopterin (MPT) converting factor; subunit 2; molybdopterin biosynthesis MoeA; oligopeptide transport system permease protein; peptidyl-tRNA hydrolase; phosphoribosyl aporibitylcytlanidine cyclo-ligase; photosystem II P680 chlorophyll a aporibityl; polyA polymerase; endonuclease; replicative DNA helicase; succinate dehydrogenase iron-sulfur protein; tRNA-Cys; tRNA-fMet(exon1); tRNA-fMet(exon2); Ycf34.

SOURCE *Synechocystis* sp. (strain:PCC6803) DNA.

ORGANISM Bacteria; Cyanobacteria; Chroococcales; *Synechocystis*.

REFERENCE 1 (bases 1 to 113064)

AUTHORS Kaneko,T., Tanaka,A., Sato,S., Kotani,H., Sazuka,T., Miyajima,N., Sugiyama,M. and Tabata,S.

TITLE Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. I. Sequence features in the 1 Mb region from map positions 648 to 928 of the genome

JOURNAL DNA Res. 2 (4), 153-166 (1995)

MEDLINE 96127529

REFERENCE 2 (bases 1 to 113064)

AUTHORS Kaneko,T., Sato,S., Kotani,H., Tanaka,A., Asamizu,E., Nakamura,Y., Miyajima,N., Hirasawa,M., Sugiyama,M., Sasamoto,S., Kimura,T., Hosouchi,T., Matsuno,A., Muraki,A., Nakazaki,N., Nairuo,K., Okumura,S., Shimpou,S., Takeuchi,C., Wada,T., Watanabe,A., Yamada,M., Yasuda,M. and Tabata,S.

TITLE Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions

JOURNAL DNA Res. 3 (3), 109-136 (1996)

MEDLINE 97061201

REFERENCE 3 (bases 1 to 113064)

AUTHORS Tabata,S.

TITLE Direct Submission

JOURNAL Submitted (30-AUG-1995) to the DDBJ/EMBL/GenBank databases. Satoshi Tabata, Kazusa DNA Research Institute, Gene Structure 2; 1532-3

Yanachino, Kisarazu, Chiba 292, Japan (E-mail:tabata@kazusa.or.jp, Tel:0438-52-3933, Fax:0438-52-3934)

COMMENT potential protein coding regions were assigned on the basis of similarity search of the ORFs and Genemark analysis.

FEATURES

source

1. 113064

/organism="Synechocystis sp."

/strain="PCC6803"

/db_xref="taxon:1143"

complement(1..3960)

/note="ORF_ID:s110178"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="BAI10440.1"

/db_xref="GI:1001201"

CDS

CDS

CDS

CDS

CDS

CDS

/translation="MSQQNVNKKRFEWCPRSGNSISLSDGGYLDDPNTKMGKACNPDLVS
LEDAIEPCLIVLGEPPVGKTELEKQSTETETICEYSGKVLNLRFTNLEKDLFR
DKTFSWLGSHLHLFLDSIDGEGILNVSALIGVDEKKRYSRHLRLRLCR
TLAFPSILLEALKALWESNFAITELAPLRIDVINAIVABGSSDDPLKAINORDIV
PLAIKPISLGFLNTRYRHNSQPPPOKKESTELBCGKLCCEVSKRALKETGND
IDGLIAVARIAAVTIFANRFAMVIDEGGVAEDEVPOKLCFGEESNGEFEITQ
EVIQEVLDGGLFSSRGLHRGMNAHQYAAEFAMVLTQNNIDLSKIKTLVLSASPT
KLIPOLHETAVLWASMRDYNFOEIIKTDPDVLQTDIPDSELRATIVANLMOYEE
KLIDRFQCVNRNARKLHSGIAFOLKVVYICDSKKEDTRIRIADIAECLNSELQGL
AKITLCPQSITIRVNAARISITGVQSTRIKLSALVWLPEDRODELKAYALKAIV
PRLTAELPEFQALTPSKRINLGMHRYFLESELVPLQDSELYIALDMLKTGVYCFG
HPEELEGDALILKAMEHVNVPGLVEGPTVVALQWHEOKLISHSPSLAQOQESLIG
NNOKRYALLIENLIESIAEKDEDEPQYAEFTVSMTEPIISSNDVMMLEKLDSENHETQ
KMAKIIIEEVDRODKOISEIKATQNNISLSDVSYFTPELDESEVTDLRHYL
RNANNNKPPDLPDPKERYIDCLEDSGDSAMWLKREMTLLPOSKRYGHNLIKLD
LTQDPKKEDEQATTKRKRIIOGATQYIOQONTDYSWIGNTNYSFSLSYHSPDEPLTIK
EAPRTLDNTSTDMKRWPTITLSVPSNFSIDSNRKCRLKYSLSYHSPDEPLTIK
LIDKENDNNIEVIDRLEMCWDEKLENNLTERKYKDSQLRNFRKLEKLEIGAGVQ
KDIALSTIRNTNSSEFDESEROKALFALKALITYSDPSNMVWIEKITYADVQGEARE
LAASCSLGSKFPLTEKOIADLYAMVWQPHYEDPDHREVMAYIVSPDNTAFPRN
NSISOLDKGTTEACAEERLQELPMLPKETLYAQOONMRETRWPEPKLLEF
LIIPESNDLSNLSITDKRTKMEDEPKITNNISDSTINAPVGSVGTNNVTCSS
PEKOKRNMWELTGVIGITVATVVGIFPNGLFNDVLKDSNGSKIEQAMPKN"
complement(4106..4678)
/note="ORF_ID:s110888"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAI10441.1"
/db_xref="GI:1001202"
/translation="MLALQPRFKSFDEYLAEDDTQLYELNGELLAVPESGLNQ
IAPFLKFAEIVGVRVGRGHLELVHGEPRKPRFPDITIRREHIAQLAKRNTIRLS
MLPELIVAVVSPGELQREBDYIAKROQYACGPEWIVDPQTQVLYVELLAGCY
EVGNFKYVMAIOSPTFPOLSLTVNEIFKDL"
4733..4930
/note="ORF_ID:ssr1513"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAI10442.1"
/db_xref="GI:1673308"
/translation="MAIAKEFEFPLQKSGRRLYITGFAPKESDITTVFPLQALTVGEI
AAEYARELFADNNTDCLYPR"
complement(4987..6327)
/note="ORF_ID:s110887"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAI10443.1"
/db_xref="GI:1001203"
/translation="MLDIOALADHAHOGAALGIKQYDIYSSVDAGVEVDSGEPRQ
VEASNRASVIRVNRKSGLVGTVSTTDIDQGISLALETAAEASAFGITDNIIDFSE
AOVPIGAGENDVAPRPMGDIITTLVYRAEDOLLSHRAITGVPGNGISDQTEEFYIN
SAGAKRYECSRYSIVYLSRAEQSGKRKRSAGEMKISBDLASLIDICDLOTVKETS
HLDYRPIPTGKTYVFSARAPLSLGAIVSNYNQSLIDKQSLAKETLGTALASSEL
GSDDALHPGNISAREPDEGTPTRRDVILKNGVLTGLHSAGAKRFRDQPTGHAN
GSKITVGAHFYHVVSEKTAESYSLDOENVVYIDKLOALHAGVNSLOGSSELPFDWM
LVNKGKSIDAIVYAGDLFVLTIVYLEBEAVTGTGGVPLVWVGLAITSB"
complement(6412..7251)
/note="ORF_ID:s110886"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAI10444.1"
/db_xref="GI:1001204"
/translation="WLSLRLRYFALMLPSLWGLYVGLILFFSPVWATESPITPLTDE
QLEVGDSLTKDAFAATEAGDEVTAKEKYTELEIEFPDPVAVWSKNGSRVSQNLDA
IADFNQATELAEQOTDYPILNGRTALEAGKEFKAIIDPNRYLAVNPEDAVYNNRGA
EGGLGNMSTALEDFQOATAIAPNFAQANTALALYELGQTEAIOGMRRLVRYKYPME
ADMRAALTAIVLWDQGEAESNMWAAVGVDSRQDLDWVKQIRRMPPSVISALDKEL
SLS"
7468..8031
/note="ORF_ID:s110895"


```
357 ycyValIleLysGIuThrLeuArgLeuAsnProValIleProGlyIleP 374
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
63074 CCAAGAGTACAGAGAAAGTATGCGGTATTCCTCTGCGGGGGGTT 63123
374 heArgValAlaLeuLysThrPheGluLeuAsnGlyTyrGlnIleProLys 390
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
63124 TTCGAGAAATTAATCCAAAGATTGCAATTCAGGGTTTCATTTCACCA 63173
391 GlyTPAsnValIleTyrSerIleCysAspThrHisAspValAlaGluI 407
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
63174 GCGTGGTATGTTCTTATCAATTAATACCAACCCAGCCGATCTGATT 63223
407 epeThrAsnLysGluGluPheAsnProAspArgpe.....S 420
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
63224 ATATCCCGACCCAGAAAGTTGATCCAGAACGGTTACTCTGATGTA 63273
420 eAlaProHisProGluAspAlaSerArgPheSerPheIleProPheGly 436
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
63274 GTGCACCTCATATCCCTCC.....TTTGCCCATGTTCCCTTTGGC 63314
437 GLYGLYLeuArgSerCysValGlyLysGluPheAlaLysIleLeuLeu 453
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
63315 GGTGGCTTAGGGAAATTTAGTAAGAAATTTGCTCGCGTGGAAATGA 63364
453 sIlePheThrValGluLeuAlaArgHisCysAspTTPGlnLeuLeuAsn 470
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
63365 ATATATTGCCACCCGATTAAATCCAAATTTGATGGACATTTGTTACCG 63414
470 lYrProP.....ThrMetLysThrSerProThrValTyrProValAspAsn 485
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
63415 GGCAAATTTAGAGTGTGTTTACCTTAGTCCCGCCGCCCAAGATTAAT 63464
486 Leu 486
|||||
63465 TTG 63467

seq_name: gb_ov:AF185266

seq_documentation_block:
LOCUS AF185266 322 bp mRNA VRT 27-SEP-1999
DEFINITION Gallus gallus CYP26 (CYP26) mRNA, partial cds.
ACCESSION AF185266
VERSION AF185266.1 GI:5924309
KEYWORDS
SOURCE
ORGANISM
. chicken.
Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
1 (bases 1 to 322)
REFERENCE
AUTHORS
TITLE
DIRECT SUBMISSION
SUBMITTED (09-SEP-1999) Cell and Molecular Biology, Tulane
University, 2000 Stern Hall 6400 Freret St., New Orleans, LA 70118,
USA
FEATURES
source
1..322
/organism="Gallus gallus"
/db_xref="taxon:9031"
<1..>322
/feature="CYP26"
<1..>322
/feature="CYP26"
/feature="contains heme binding domain; RA-metabolizing
enzyme"
/codon_start=2
/product="CYP26"
/protein_id="AA056546.1"
/db_xref="GI:5924310"
/translation="GFRVALKLELNGCYIPKGMNVLYSTCDTHVADLFTDKDENP
DREMSPEDESSRFSFIPFGGLRSCVGEFAVLLKIFVELARSCDWQLNGPPTM
KTSPT"
```

```
BASE COUNT 74 a 82 c 85 g 81 t
ORIGIN
alignment_scores:
Quality: 535.00 Length: 107
Ratio: 5.144 Gaps: 0
Percent Similarity: 97.196 Percent Identity: 89.720

alignment_block:
us-09-668-482-4 x AF185266 ..
Align seg 1/1 to: AF185266 from: 1 to: 322

373 GlyPheArgValAlaLeuLysThrPheGluLeuAsnGlyTyrGlnIleP 389
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2 GGGTTTCGGGTTGCTCTCAAGACCTTGAGCTGAATGTTACAGATCCC 51
389 oLysGlyTPAsnValIleTyrSerIleCysAspThrHisAspValAlaG 406
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
52 CAAGAGCTGGAATGTTATTATACAGATATCTGCATACCCAGATGGCGG 101
406 lIlePheThrAsnLysGluGluPheAsnProAspArgPheSerAlaPro 422
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
102 ATCTCTTACCGCAAGAGAGAGATTCAACCCGACCGCTTCATGTCTCG 151
423 HisProGluAspAlaSerArgPheSerPheIleProPheGlyGlyLe 439
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
152 TCYCCGAGAGACTCTGTAGGTTCAGTTCAATTCCTTTGGTGGGGCTT 201
439 uArgSerCysValGlyLysGluPheAlaLysIleLeuLysIlePheT 456
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
202 GAGAGACTGGGTAGGCAAGAGTTTCCAAAGTCTTCAAAATATTTA 251
456 hValGluLeuAlaArgHisCysAspTTPGlnLeuLeuAsnGlyProP 472
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
252 CGGTGAGAGCTGGCGGAGCTGTGACTGGCAGCTTAATGAGACCTCG 301
473 ThrMetLysThrSerProThr 479
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
302 ACAATGAAACCAAGCCCACT 322

seq_name: gb_p12:AF318211

seq_documentation_block:
LOCUS AF318211 1494 bp mRNA PLN 15-FEB-2001
DEFINITION Taxus cuspidata 5-alpha-taxadienol-10-beta-hydroxylase mRNA,
complete cds.
ACCESSION AF318211
VERSION AF318211.1 GI:12656591
KEYWORDS
SOURCE
ORGANISM
. Taxus cuspidata.
Taxus cuspidata
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Coniferopsida; Coniferales; Taxaceae; Taxus.
1 (bases 1 to 1494)
REFERENCE
AUTHORS
TITLE
DIRECT SUBMISSION
SUBMITTED (30-OCT-2000) Institute of Biological Chemistry,
Washington State University, Pullman, WA 99164, USA
FEATURES
source
1..1494
/organism="Taxus cuspidata"
/db_xref="taxon:99806"
1..1494
/feature="cytochrome P450-like protein"
/codon_start=1
```

```

/product="5-alpha-taxadienol-10-beta-hydroxylase"
/protein_id="AAK00946.1"
/db_xref="GI:12656592"
/translation="MDSFTPLRSISTKRGQLESPALISLTAPLAILLILFLRYNH
KSSVLPKGLFGLIGETIOLRLRSETPOKRFDDRLKRPYMTSLIGHPTVVL
CGPAGNKLVLSENEEDLVEMEGERSFMKLIGEDSIYAKGEDHRIILRLARLGAOL
QNYLHSGSEIGHNENKMGDEKVLPLVAGLSLSTLFPVNGHOOKOLHL
LETLIVGLSLPDEPGTRYRGLQARLKLDELISLJKRRRLRSGLASDDOLLS
VLFLFRDKGSLNDQGLIDNPSAMFASDYDTVA PMALIKRLXSNBEYHEKVOEO
LEITGNKEGSEISKDKLSMKTQANQVESIRMPYVGLFKRAITDIIDGYTIPK
GMRVLCSPYTHHREYRPEEPERPSRFEDEGRHVTYVYVPEGGLRTPCGWBFK
LEILFVHHEKFNSSYIPVDNKEVLSDDLPLPLPANGFSIKLFPSR"
BASE COUNT      429 a      315 c      314 g      436 t
ORIGIN

```

```

alignment_scores:
  Quality: 524.50      Length: 496
  Ratio: 1.720        Gaps: 11
  Percent Similarity: 61.492      Percent Identity: 28.427

```

alignment block:

US-09-668-482-4 x AF318211 ..

Align seg 1/1 to: AF318211 from: 1 to: 1494

```

4 ProAlaLeuAlaSerAlaLeuCysThrPheValLeuProLeuLeu 20
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
61 CCTGATATCTTCCCTTACCCGCGACCTATTCGCCATTAATTTCTTCT 110
20 ubHeuAlaAlaAlaLeuLeuTrpAspLeuTrpCysValSerGlyArg 37
|:|||||
111 CTGGCTCTCCGCTTAC.....A 127
37 spArgSerCysAlaLeuProLeuProGlyThrMetGlyPheProPhe 53
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
128 ATCACCAGATCCTCTGTAACTTCCCTCGGAGAAAGTTAGCTTTCTCTTC 177
54 PheGlyGluThrLeuGlnMetValLeuGlnArg..... 65
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
178 ATCCGGGAGACCATCAATATTATTCGCGACCTCCGTCAGAAACCTCA 227
66 .LysPheLeuGlnMetLysArgArgLysArgLysPheLeuLeuTrp 82
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
228 AAAGTTTGTGATGATGATTAAGAAATTCGCTCTGTTACATGACTT 277
82 isLeuPheGlyArgProThrValArgValMetGlyAlaAspAsnValArg 98
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
278 CCCATATTGGGCATCCACAGTTTACTCTCGGGCCTCGGGGAAACAA 327
99 ArgIleLeuLeuGlyAspAspArgLeuValSerValHisTrpProAla 115
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
328 TTAGTCTCTTTCGACGAGACAAAGCTGTAGAGTGAAGGCGCCAGATC 377
115 rValArgThrIleLeuGlySerGlyCysLeuSerAsnLeuHisAspSer 132
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
378 TTTCATGAACCTGATGGGGAAGATTCATGTTCTTAAGAGCGAG 427
132 erHisLysGlnArgLysValIleMetArgAlaPheSerArgGlnAla 148
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
428 ATCATTCGCGATTTACGCACTGCTGCTGCTTTTGGGCGCTCAAGCT 477
149 LeuGluCysTrpValProValIleThrGluGluValGlySerSerLeu 164
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
478 TTACAATAATTTCTGGTAGATGATTCAGAAATAGGACACCATTTTCAA 527
165 .GluGlnTrpLeuSerCysGlyGluArgGlyLeuLeuValTrpProGlu 181
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
528 TGAAGAAATGGAG.....GCTAAAGATGAAGTGAAGTCTCTCTTGG 571
181 alLysArgLeuMetPheArgIleAlaMetArgIleLeuLeuGlyCysGlu 197
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
572 TAAAGGGCTTATCTTCTCCATTCGACACACCTGTTTTCGAT..... 615

```

```

198 ProGlnLeuAlaGlyAspGlyAspSerGluGlnGlnLeuValGlnAlaPhe 214
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
616 .....GTAAATGATGACACACACACAGACCATTCATCTTCTCT 656
214 eGluGlnMetThrArgAsnLeuPheSerLeuProIleAspValProPhe 231
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
657 GGAACATATCTCTGCGGAAGTTTGTCAGTCCGCTGACCTTCACGAA 706
231 eArgLeuTrpArgGlyMetLysAlaArgAsnLeuIleHisAlaArgIle 247
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
707 CTGATTCGTAAAGGCTTCAGCGCGCGCTGAAGCTTGATGAATCTCTC 756
248 GluGlnAsnIleArgAlaLysIleCysGlyLeuArgAlaSerGluAla 264
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
757 TCCTCTTAATTAACGACGAGAGAGATCTGCTTACAGCATAGCTTC 806
264 yGlnGlyCysLysAspAlaLeuGlnLeuLeuIleGluHisSerTrpGlu 281
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
807 TGATGAT...CAAGATCTACTGTCGGTGTCTACCTTACAGATGAA 853
281 rglGlyGluArgLeuAspMetGlnAlaLeuLysGlnSerSerThrGlu 297
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
854 AAGGAACCTCAGCTGACAGACAGGAGATCTGACCAACTTTCTGCTATG 903
298 LeuPheGlyGlyHisGluThrThrAlaSerAlaIleThrSerLeuIle 314
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
904 TTTCATGCTCATATGACACCACTGTTGACCAATGGCCTTGATATTA 953
314 rTyrlLeuGlyLeuTrpProHisValLeuGlnLysVal...ArgGluGlu 330
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
954 GCTTCTATATCCTCAATCTGAATACCATGAAAGATTTTCAAGACACT 1003
330 euLysSerLysGlyLeuLeuCysLysSerAsnGlnAspAsnLysLeu 346
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1004 TGGA.....ATAATTGGCAATAAAGAGAGAGAGAAATCACT 1044
347 MetGluIleLeuGlnGlnLeuLysTrpIleGlyCysValIleLysGlu 363
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1045 TGGAGGATTTGAAATCTGTAATATATACATGCAAGCAAGTTCAAGAAC 1094
363 rLeuArgLeuAsnProProValProGlyGlyPheArgValAlaLeuLys 380
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1095 ACTAGCAATGTACCCACAGTTTGGAAATATTTGTAAGCTATACATCG 1144
380 hrPheGluLeuAsnGlyTrpGlnIleProLysGlyTrpAsnValIleTr 396
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1145 ATATTCATATGATGGGTATACAAATTCAAAGAGATGGAGGTTTATGT 1194
397 SerIleCysAspThrHisAspValAlaGluIlePheThrAsnLysGlu 413
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1195 TCACCTTACTACACATCTGAGAGAGATCTTCCCGACCTCGAAGAA 1244
413 ubPheAsnProAspArgPheSerAlaProHisProGluAspAlaSerArg 429
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1245 ATTCAGGCTTCAAGATTT.....GAGATGAAGGACAGGC 1279
430 .....PheSerPheIleProPheGlyGlyGlyLeuArgSerCys 442
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1280 ATGTACTCTTACACATATGATGACATTTGGAGAGGCTCGGCACATGT 1329
443 ValGlyLysGluPheAlaLysIleLeuLeuLysIlePheThrValGlu 459
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1330 CCAGATGGGAATTTCAAGATGAATATATACATGTTTGCACATATT 1379
459 uAlaArgHisCysAspTrpGlnLeuLeuAsnGlyProProThrMetLys 476
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1380 CGTTAAATTTCAACACACTCATCTCCAGTTGATCCCAATGAAAAAGTTT 1429
476 hrSerProThrValTrpProValAspAsnLeuProAla 488
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1430 TATCAGAT.....CCACTACCTCTCTCCCTGCC 1458

```

seq_name: gb.p11:AC006931

seq_documentation_block:
 LOCUS AC006931 116205 bp DNA PLN 05-APR-2000
 DEFINITION Arabidopsis thaliana chromosome II section 231 of 255 of the
 complete sequence. Sequence from clones F14N22, F7D19.
 AC006931 AE02093
 VERSION AC006931.5 GI:6598641
 KEYWORDS HTG.
 SOURCE thale cress.
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 Brassicales; Brassicaceae; Arabidopsids.
 1 (bases 1 to 116205)
 Lin.X., Kaul,S., Rounsley,S.D., Shea,T.P., Benito,M.-I., Town,C.D.,
 Feldjuyum,T.V., Buell,C.R., Bowman,C.L., Barnstead,M.E.,
 Koo,H., Moffat,K.S., Cronin,L.A., Shen,M., Vanden,S.E., Umayam,L.,
 Tallon,L.J., Gill,J.E., Adams,M.D., Carrera,A.O., Greasy,T.H.,
 Goodman,H.M., Somerville,C.R., Coppenhaver,G.P., Preuss,D.,
 Nierman,W.C., White,O., Eisen,J.A., Salzberg,S.L., Fraser,C.M. and
 Venter,J.C.

TITLE Sequence and analysis of chromosome 2 of the plant Arabidopsis
 thaliana
 JOURNAL Nature 402 (6763), 761-768 (1999)
 MEDLINE 20083487
 PUBMED 10617197
 REFERENCE 2 (bases 1 to 116205)
 AUTHORS Lin,X.
 COMMENT Direct Submission
 Submitted (09-MAR-2000) The Institute for Genomic Research, 9712
 Medical Center Dr., Rockville, MD 20850, USA
 On Dec 17, 1999 this sequence version replaced gi:4512656.
 The sequence and annotation of chromosome 2 were merged from those
 of the individual clones on this chromosome after removing
 overlaps. For detailed information, please see the TIGR web site
 (http://www.tigr.org/tldb/at/at.html).

Genes were identified by a combination of three methods: Gene
 prediction programs including GRAIL
 (http://arthur.epm.ornl.gov/pub/xgrail), Genefinder (Phil Green,
 University of Washington), Genscan (Chris Burge,
 http://genomic.stanford.edu/GENSCAN.html), and NetplantGene
 (http://www.cbs.dtu.dk/services/NetGene2/), searches of the
 complete sequence against a peptide database and plant EST
 databases at TIGR, and manual curations based on those analyses.
 Annotated genes are named to indicate the level of evidence for
 their annotation. Genes with similarity to other proteins are named
 after the database hits. Genes without significant peptide
 similarity but with EST similarity are named as 'unknown' proteins.
 Genes without protein or EST similarity, that are predicted by two
 or more gene prediction programs over most of their length are
 annotated as 'hypothetical' proteins. Genes encoding tRNAs are
 predicted by tRNAscan-SE (Sean Eddy,
 http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats were
 identified by RepeatMasker (Arian Smit,
 http://ftp.genome.washington.edu/RM/RepeatMasker.html). Genes are
 numbered from the top to bottom of the chromosome.

We thank the CSHL/MashU/ABI consortium for sequencing BAC clones
 FBP23, F5J6, T17A5, and T13L16, the ESSA group for sequencing clone
 F13D4, and Scott Jackson, Jiming Jiang, Klaus Meyer, Eric Richards
 and Satoshi Tabata for helpful assistance. In addition, we would
 like to thank the TIGR Bioinformatics Department, especially Lixin
 Zhou, Hanif Khalak, Michael E. Heaney, Lily Fu, Feng Liang, Jeremy
 Peterson, Michael Holmes, and Delwood Richardson for software and
 database support.

This work was supported by the National Science Foundation,
 Department of Energy and the US Department of Agriculture.

Address all correspondence to: at@tigr.org.

FEATURES	Location/Qualifiers
source	1..116205 /organism="Arabidopsis thaliana" /cultivar="Columbia" /db_xref="taxon:3702" /chromosome="II" complement(<1..2131) /note="Sequence from clone F14N22"
misc_feature	<1226..>3344 /gene="At2g42650"
mRNA	<1226..>3344 /gene="At2g42650"
gene	<1226..>3344 /gene="At2g42650" /note="F14N22.8; predicted by genscan"
CDS	1226..2344 /gene="At2g42650" /note="hypothetical protein" /codon_start=1 /protein_id="AA18664.1" /db_xref="GI:6598642" /translation="MSRVSPKTVDAVKALVKEGNEKSRTEKPOLLEEDGFYLVVAL KTIPOREPMTNAYRIPIPLPIINTEDSEPELCIIDDEPESGLEEDAKKIKSENIPI TKVYKLSKLSNDYGSFESKSKLSDSYMPESDRRIYIMPLIKIGKPFOSKYPVAD LKHNMRKQLEKACGAAMFFRTGSCSAIVAKLSMESDVIYENVATINGVDYVPS RMKYIRSLHLKLSLSLPLVQYVYQLKIDPLGVEEVNKGSLKSDVDSKSSV KTKKKNCKIHEVRYMDSNVSETLGDDEDFSDVEVADVLNAGDKKKRKKMSKSSK AVSGKPIVSKNGCKSKIKLIDIDEGSGGFFKAKTKRR" complement(2132..>116205) /note="Sequence from clone F7D19" complement(join(<2761..2943,3035..3164,3258..3372, 3450..3489,3570..3646,3732..>3954)) /gene="At2g42660" complement(<2761..>3954) /note="F7D19.34; predicted by genscan" complement(join(2761..2943,3035..3164,3258..3372, 3450..3489,3570..3646,3732..>3954)) /gene="At2g42660" /note="hypothetical protein" /codon_start=1 /protein_id="AAD21740.1" /db_xref="GI:4512686" /translation="WAGRSSPGKSTWIDLNRETVKKNNEEESVDEDOITNNVRO YIRSMPEPLRWTPDLHSFVAVQRLGPPDAPTRKLYLEMNKLGSITAHYKSHLOMT RSKLLEPSSRFEGAFMGSGYLMADMSKICPHSDLRHAYNKTSPVRLNDVAV TNLGNFLMRPSSWFSGICRNDRDSTENKTPLEIRKKNNEKRVDEEVSVKRLKS MGGGICLPERRNCQRKPTDINTLSLSLSTSSB"
repeat_region	3681..3724 /rpt_family="(GA)n"
repeat_region	4087..4204 /rpt_family="(TA)n"
repeat_region	5756..5806 /rpt_family="(TA)n"
repeat_region	8844..8989,9284..>9396) complement(join(<7700..7818,7906..8112,8195..8348, 8844..8989,9284..>9396)) /gene="At2g42670"
gene	complement(<7700..>9396) /gene="At2g42670"
CDS	/note="F7D19.33; predicted by genscan" complement(join(7700..7818,7906..8112,8195..8348, 8844..8989,9284..>9396)) /gene="At2g42670"
repeat_region	/note="hypothetical protein" /codon_start=1 /protein_id="AAD21739.1"
repeat_region	/db_xref="GI:4512685"
repeat_region	/translation="MPYFAQRLYNNCKASFSSDGPITDALKYRNVLKIKPSDVG EDDAQLARSRSPLMERKNSGNSPAIRYLIHEDDSRIGIFCPMPSMPLNNHPS MVLKLVYGSNHSKSIDMLEPULTEPDPSEARPAIVYDTEMTASPYTLTXPS GGNHCFKATITCALIDLIAPRYSSSEHRRHCTYFKRSRRDPLGLEVDGEVYDVYV LEEFQPPDFVIRCAIPIRYRGPVIRT"
repeat_region	10075..10153 /rpt_family="(GAA)n"
mRNA	join(10610..10809,11165..11248,11522..11604,11693..11994)

```

gene      /gene="At2g42680"
           10610..11994
           /gene="At2g42680"
           /note="F7D19.32"
CDS       join(10668..10809,11165..11248,11522..11604,11693..11812)
           /gene="At2g42680"
           /note="unknown protein"
           /codon_start=1
           /protein_id="AAD21738.1"
           /db_xref="GI:4512684"
           /translation="MAGIGPITODEPVIYIRKKPANAAKRDERTVNARRSGADIER
VRKRNAGNKAASSTSLNTKMDLDDPENLTHEVPTBLKALIMQARIDKLTQSOLA
QIINEKPOVIOEYSGAIEPMQILSKLERLAGALRCK"
mRNA      complement(join(11994..12273,12362..>13378))
           /gene="At2g42690"
           complement(11994..>13378)
           /note="F7D19.31"
           /gene="At2g42690"
           complement(join(12052..12273,12362..13378))
           /gene="At2g42690"
           /codon_start=1
           /product="putative lipase"
           /protein_id="AAD21737.1"
           /db_xref="GI:4512683"
           /translation="MATTTISWEELLSKMDTILDPIDSLREILIRCGDFQATYD
AFVNDNSKTCGASRYGKSFEDKVMLENAGVNFLYATARYSLPEGILLOSQR
DSMDRESNMFICYIAVTSDESKALGRREIYALRGSTNRYEMVNLGAPRTSADPLH
GPEODSGGVGVEGTTFDSEDEEGCKMLGMLITYSNBESKFTKLSRDLAKI
KELLKTKDEKPSIVLGHSLGATEAVLAADIAENGSSDDVPYTAIFYEGPOYKNE
FRDVMGHKNTKILHVRNTIDLTRYPGILGYVDIGINPYIDIKKSFSDSRNPGD
MHNQAMLIHYVAGNRGKGEKRLMVKRSIALVNSCEFLNAECIVPSGWVKNKGLI
KNIEGVEVLAVEEPEYEP"
mRNA      join(<14160..14207,14422..14565,14779..14890,15255..15434,
           15513..15664,15832..17364,17446..17514,17602..17786)
           /gene="At2g42700"
           <14160..17786
           /note="F7D19.30"
           /gene="At2g42700"
           join(14160..14207,14422..14565,14779..14890,15255..15434,
           15513..15664,15832..17364,17446..17514,17602..17786)
           /gene="At2g42700"
           /note="unknown protein"
           /codon_start=1
           /protein_id="AAD21736.1"
           /db_xref="GI:4512682"
           /translation="MALIDVAISCLNSIREIEEDVKDAIYIDAGCTESFQVGFAPL
FLETGARAVCSLEMTSLDAVDWNSKSDCAKRIYIMTSRLNDAHRYMLCLSTHEG
PDARREYETLLVODYNEHTKSKDKISKDKGKSSALSLSTMEPTSENVDISGGA
OGSAVEASLSROHEDSLSPGLPISTGMSPTDDVPPGATLTAAFLYIOLAKMKLKE
IFSIGDQSKNVGKLTIDMSSYYDVARRKRSGLLIDRLITPCCHGDSLDFRIFS
SLPRAERFSOALKGQVPSINRPSLDVQVPLGELLNEBPSKIDSGLPESIEAFIRG
WDSTSAPOVNGLENECDKSTNMTELLNGSLAVTFRCTPLEMIDIRKTDGAV
LVKMWLEALRENIYINVARAPGYATPELOAIMIKALSOSLSLKNKGIIIOGAAT
AAALDESQAKMDFSSAEMMLINYSAGDTSGGLAOSLIDINKSAVLEAVLAKKNEKD
SSSGSLSPFDALLITVGYIAGENPDSGSGSPFMOEHFELIVAVLEPNPA
GNIKFTNGLELEELGRNLKSEPTKEIIPSDDOLDLALDDDDPKGKGDEEEVYDVS
KADESIDMQLKLDLRDVSLEFRLKLSLRLRNLPLRGSLASESSPEPESGNK

```

alignment_scores:
 Quality: 468.00 Length: 526
 Ratio: 1.683 Gaps: 14
 Percent Similarity: 52.852 Percent Identity: 26.046

alignment_block:
 US-09-668-482-4 x AC006931 ..

Align seg 1/1 to: AC006931 from: 1 to: 116205

```

44 LeuProProGlyThrMetGlyPheProPheNegIyGluThrLeuGlnMe 60
   ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
67490 CTTCTCTCGTGGAGAAATGGGCTCCATGAGATTGGAGAAATGATTT 67539
60 tValLeuGlnArgArgLys.....PheLeuGlnMetL 71

```

```

67540 CTACAAAGCTCAGAAAAGCAACAGATTTTGGAGATTGTTGTAATCCAA 67589
       ::::::::::: ||:::::
71 ysArgArgLysIyrgIyPheIleTyIyLysThrHisLeuPheNegIyArgPro 87
   :: ||::::||| ||::::||| ||::::||| ||::::||| ||
67590 GGAATCATCAACATGAGGACATCTTCCAGAACAAAGAAATGAGATGACCA 67639
88 ThValArgValMetGlyAlaAspAsnValArgArgIleLeuLeuGlyAs 104
   ||:: ||| ||||| ||| ||||| ||| ||||| |||
67640 ACATATGATATCAATGAGAGCTGAGAACCAATAGGCTCATATTATCCAAAG 67689
104 pasPAArgLeuValSerValHisTrrProAlaSerValArgThIleLeuG 121
   : ||||| ||||| ||||| ||||| ||||| |||||
67690 GTTTAGCTTGCTGGTGGAGCTCGGCTTCATCTCTGTTCAACTAATGG 67739
121 IySerGlyCysLeuSerAsnLeuHisAspSerSerHisLysGlnArgLys 137
   ||::::|||::::: ||::::: ||::::: ||::::: ||:::::
67740 GCATGAACTGCATCATGGCTAAACAAAGGCGAAGCATCGGGCTTCAGA 67789
138 LysValIleMetArgAlaPheSerArgValAlaLeuGluCysTyrValPr 154
   ::::: ::::::||| ::::::||| ::::::||| ::::::|||
67790 GGGATTGTACTATATAGTCTTACTCATATTGGGTTAGAGCTTAATATAC 67839
154 oValIleThrGluGluValGlySerSerLeuGlu...GlaTrpLeuSerC 170
   | :::: ||| ||| ||| ||| ||| ||| |||
67840 TAACTCTGTGATACCGTTAAGTTTCATCATGAAACCGAATGGCGA.... 67885
170 ysgIyGluArgGlyLeuLeuValTyrProGluValLysArgLeuMetPhe 186
   ||::::|||::::: ||::::: ||::::: ||::::: ||:::::
67886 ..GGCAAGGAGAGATTAGCTTACCGCTTGGCAAAAGTTTCTCACTTTC 67933
187 ArgIleAlaMetArgIleLeuLeuGlyCysGluProGluLeuAlaGlyAs 203
   ::::::|||::::: || ||| ::::: ||::::: ||:::::
67934 ACTGTAGTATTTCAGTGTGTGTATGGAGATTAAGTGAGATTTGG.... 67978
203 pglYAspSerGluGluGlnLeuValGluAlaPheGluGluMetThrArgA 220
   ::::::|||::::: ||::::: ||::::: ||::::: ||:::::
67979 .....ATGTTAAGGCTTTTGAGAGGCTTTTGAGAG 68009
220 snLeuPheSerLeuProIleAspValProPheSerGlyLeuTyrArgGly 236
   ::::::|||::::: ||| ||| ||| ||| ||| |||
68010 GCTTTTCCGTTTGGCCGGTTCGAGTTTCTCTTCTTAAGTTTCTTAACGA 68059
237 MetLysAlaArgAsnLeuIleHisAlaArgIleGluGlnAsnIleArgAl 253
   ||||| ||| ||| ||| ||| ||| ||| |||
68060 AAGAAAGCAAGCTAGATGAGACATTTCTTGGGAGAGGTCAGGGA 68109
253 alysiIleCysGlyLeuArgAlaSerGluAlaGlyGlnGlyCysLysAspA 270
   ||| ||| ||| ||| ||| ||| ||| |||
68110 GAAG.....CGAAGAGAAATGGAAAAAGGAGGAGCTGAGAACCC 68147
270 lalLeuGlnLeuLeuIleGluHisSerTrpGluArgGlyGluArgLeuAsp 286
   ||:::: ||:::: ||:::: ||:::: ||:::: ||::::
68148 CAACACCAACGCTTTC.....TCGAGCGTAAAGA 68176
287 MetGlnAlaLeuLysGlnSerSerThrGlu..... 296
   :::: ||:: ||:: ||:: ||:: ||:: ||::
68177 GAAGAGTGTATCAAAAGGTGATTAACGAAAGAGGTTGATGATTAAT 68226
297 .....leuLeuPheGlyGlyHisGluThrThrAlaSerAlaIalaThs 311
   ||::::|||::::: ||::::||| ||::::||| ||::::|||
68227 GGTTTTGGTATGTTTGGAGCTCATGACCAACCTTCTTACGCCATGTCCA 68276
311 erLeuIleThrTyLeuGlyLeuTyPro.....HisValLeuGlnLys 320
   :::: ||:::: ||:::: ||:::: ||:::: ||::::
68277 TGACTTTCAAGAATGTTAGCTCAGACCAACCACTTGCTGACACTGCTCTT 68326
321 .....HisValLeuGlnLys 325
68327 CAAGTAACAATCAACTGATATCAGAAATGTTCTTCTTATCAGCAAAA 68376
325 sValArgGluGluLeuLys.....SerLysGlyLeuL 336
   | :::::: ||::::

```



```

150 lucstyValProvalIleThrGluValGlySerSerIleuGln 166
   :: |||::||| ||| |||::|||:: |||::|||::|||
487 CTCTATATACCGTACATGAGAAACGTTATTACTGTCGGAAG 536
   |||::|||::|||::|||::|||::|||::|||::|||
167 TrpLeuSerCysGlyLysGlyLeuValTyrProGluValLysar 183
   |||::|||::|||::|||::|||::|||::|||::|||
537 TGAGCCAAATGGAGGTTGAGTCTTCTGACATC.....CTGCCTAA 580
   |||::|||::|||::|||::|||::|||::|||::|||
183 GLeuMetPheArgIleAlaMetArgIleLeuGlyCysGluProGln 200
   |||::|||::|||::|||::|||::|||::|||::|||
581 GCCACACTTAGAGTATCATCTGACATCTTCTGACGCTGAGAGTGA 630
   |||::|||::|||::|||::|||::|||::|||::|||
200 euAlaGlyAspGlyAspSerGluGlnIleuValGluAlaPheGlu 216
   :: |||::|||::|||::|||::|||::|||::|||
631 ATGTATGATGCA.....TTGAGAGGAAATATACAGCT 665
   |||::|||::|||::|||::|||::|||::|||
217 MetThrArgAsnLeuPheSerLeuProIleAspValProPheSerGly 233
   |||::|||::|||::|||::|||::|||::|||
666 CTTAATGATGGGTTGAGCAATGGCAGTCAATATTCCTGATTTGCTTA 715
   |||::|||::|||::|||::|||::|||::|||
233 CTATATGAGTATGAGTATGAGTATGAGTATGAGTATGAGTATGAGTAT 250
   |||::|||::|||::|||::|||::|||::|||
716 TCATAGAGAGCTCAAGGCTAGAGAAACCTTGTGCTGCTT...CAGT 762
   |||::|||::|||::|||::|||::|||::|||
250 snIleArgAlaLysIleCysGlyLeuArgAlaSerGluAlaGlyGln 266
   |||::|||::|||::|||::|||::|||::|||
763 CCATGATGCTAGCAGCAGCAATCAAGAGAGCAACATTTATCAAAAT 812
   |||::|||::|||::|||::|||::|||::|||
267 CysLysAspAlaLeuGlnIleuLeuIleGluHisSerTrpGluArgLys 283
   |||::|||::|||::|||::|||::|||::|||
813 AAGAAAGATATGTTAGACAATCTCTTAATGTTAAAGATGAGATGAAA 862
   |||::|||::|||::|||::|||::|||::|||
283 uArgLeuAspMetGlnAlaLeuLysGlnSerSerThrGluLeuPheG 300
   |||::|||::|||::|||::|||::|||::|||
863 AACTTGGATGAGCAAGAGATATGATGATGCTCTTGCATGATCTTAAG 912
   |||::|||::|||::|||::|||::|||::|||
300 LysGlnIleGlnIleThrAlaSerAlaIleThrSerIleuIleThrLys 316
   |||::|||::|||::|||::|||::|||::|||
913 CTGGTCAATGATCTCTGCGCACACATTAATGTGGCTACCGTTTTCTTA 962
   |||::|||::|||::|||::|||::|||::|||
317 GlyLeuTyrProHisValIleuGlnLysValArgGluGluLysSer 333
   |||::|||::|||::|||::|||::|||::|||
963 CAAGAACAACCCGAGGTTCTTCAAGAGCAAGAGCTGAA.....CA 1003
   |||::|||::|||::|||::|||::|||::|||
333 sGlyLeuLeuCysLysSerAsnGlnAspAsnLysLeuAspMetGluIle 350
   |||::|||::|||::|||::|||::|||::|||
1004 AGAAATGATCTGAAAAGTAGGCTGAAAGCTCAAAAGGCTTATCTCTAA 1053
   |||::|||::|||::|||::|||::|||::|||
350 euGlu.....GlnLeuLysTyrIleGlyCysValIleLysGluThrLeu 364
   |||::|||::|||::|||::|||::|||::|||
1054 AAGAAACCCGGAATATGAGATCTTATCGCAGGTTGTGATGAGACACTT 1103
   |||::|||::|||::|||::|||::|||::|||
365 ArgLeuAsnProProValProGlyLysPheArgValAlaLeuLysThrPh 381
   |||::|||::|||::|||::|||::|||::|||
1104 CGAGTCAATCAATCTCTCACTACGCTTTTGCAGAGCAAGAGCTAGAGT 1153
   |||::|||::|||::|||::|||::|||::|||
381 eGlnLeuAsnGlyTyrGlnIleProLysGlyTyrAsnValIleTyrSer 398
   |||::|||::|||::|||::|||::|||::|||
1154 CGAAATGATGCTATTTGATCCAAAGGTTGGAAGGTTTGACATGCT 1203
   |||::|||::|||::|||::|||::|||::|||
398 LeCysAspThrHisAspValAlaGluIlePheThrAsnLysGluGluPhe 414
   |||::|||::|||::|||::|||::|||::|||
1204 TTAGGAGATGTCACATGACCCCTGAGTCTTCCAGATCCAGAAATTTT 1253
   |||::|||::|||::|||::|||::|||::|||
415 AsnProAspArgPheSerAlaProHisProGluAspLysArgPhe 431
   |||::|||::|||::|||::|||::|||::|||
1254 GATCCTGCTAGATGAGATATGTTTGTACCAAAAGCTGCT.....GC 1297
   |||::|||::|||::|||::|||::|||::|||
431 rPheIleProPheGlyGlyLysArgSerCysValGlyLysGluPhe 448
   |||::|||::|||::|||::|||::|||::|||
1298 GTTCTTCTCTTTTGTGCTGGAAGCCATCTATGCCAGGAATGATCTCG 1347

```

```

448 lAlaLysIleLeuLysIlePheThrValGluLeuAlaArgHisCysAsp 464
   |||::|||::|||::|||::|||::|||::|||
1348 CTAGAGCTGAAATTTCAATTTTCTTCATCATCTCTCTCAAAATATCAG 1397
   |||::|||::|||::|||::|||::|||::|||
465 TrpGlnLeuLeuAsnGlyProProThrMetLysThrSerProThrVal 481
   |||::|||::|||::|||::|||::|||::|||
1398 CTGAACGAGGACAAACCCGAGATGCTGATGATGATGCTGCTACACAG 1447
   |||::|||::|||::|||::|||::|||::|||
481 rProValAspAsnLeuProAlaArgPheThrHis 492
   |||::|||::|||::|||::|||::|||::|||
1448 ACCAAGTGAATATGCTTGGCAAGAAATATGATAT 1481

```

```

seq_name: gb_p14:SLU54770
seq_documentation_block:
LOCUS      SLU54770      1395 bp      mRNA      PLN      18-OCT-1996
DEFINITION Solanum lycopersicum cytochrome P450 homolog (Dwarf) mRNA, complete cds.
ACCESSION  U54770
VERSION    U54770.1  GI:1421740
KEYWORDS   tomato.
SOURCE      Lycopersicon esculentum
ORGANISM   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
REFERENCE   1 (bases 1 to 1395)
AUTHORS    Bishop,G.J., Harrison,K. and Jones,J.D.
TITLE      The tomato Dwarf gene is isolated by heterologous transposon tagging
MEDLINE    encodes the first member of a new cytochrome P450 family
9626705
JOURNAL    2 (bases 1 to 1395)
MEDLINE    Bishop,G.
AUTHORS    Direct Submission
TITLE      Submitted (11-APR-1996) Gerard Bishop, Sainsbury Laboratory, John
JOURNAL    Innes Centre, Colney Lane, Norwich, Norfolk, NR4 7UH, UK
FEATURES
source
1..1395
Location/Qualifiers
1..1395
/organism="Lycopersicon esculentum"
/strain="GCR758"
/db_xref="taxon:4081"
1..1395
/gene="dwarf"
1..1395
/gene="dwarf"
/codon_start=1
/product="cytochrome P450 homolog"
/protein_id="AAB17070.1"
/db_xref="GI:1421741"
/translation="MAFLIFLSFGLCTALRLMNQVKNOKNLPRGTMGWPLF
GETTEFLKGPSPFNKORARVGSPEFKSLICPTIYSMSLELRILIVNAKCLIVRY
POSMIDILKCNIAAVNGSAHKYMRGALLSLIPMTRODLKRIDEFMSHLTNDN
KVIDIDKTKMAFLSKKQIAGIESISLAQERWSEFNVLTLSLPNLPTNTHR
GFOARKITVNLRLTLLEERASKELIQHDMIGYLMNSEATKFKLTDEMIDLITTIYS
GLETVASTTSMAAKYVLHDHPRKVELERKEHMAIREKKREDPIDYNDYRMRTRAVI
LETSRLATVNGVLRKRTQDMEINGYLIIPKGRILYVYTRLEINDPRLYDPYVSFNPR
WMDKSLKHONSFLVFGGGTRQCPGKELGVAIEISTPLHYFTKRWMEIGSDKLMTKPR
VEAPNGLIRIVSAH"

```

```

BASE COUNT      470 a      242 c      268 g      415 t
ORIGIN

```

```

alignment_scores:
Quality:      444.00      Length:      467
Ratio:        1.620      Gaps:        12
Percent Similarity: 58.672      Percent Identity: 27.409

```

```

alignment_block:
US-09-668-482-4 x SLU54770
Align seg 1/1 to: SLU54770 from: 1 to: 1395

```



```

/chromosome="2"
/map="section 181"
/nc="ecotype: Landsberg erecta"
1..1740
/gene="KA02"
73..1542
/gene="KA02"
/nc="cytochrome P450; CYP88A4; AtKA02"
/codon_start=1
/product="ent-kaurenoic acid hydroxylase"
/protein_id="AK11565.1"
/db_xref="GI:13021856"
/translation="MTETGLIMFPLIILGLFVLRKRVNWIYVSKLGEKKHYL
PPGDLGWVIGNMSPFLAEKTSDESPISQITRYRGRTGTYKAMMEGYPCVILVTPPE
TCRRVLDNDPAFHGMPSKTMKLGKSPFNGISFEERHRLRTRTSAPNGPEALSVYI
OFIEEYNTDLEKMSKMEITFELSLKLTETKVMITFLESSEHMDLSRETNNL
KGVAMKINLPGRVHRLKARKKLVAFOSIVNRNRQRONTSSNRKMDLNDITV
KDEGRVLNDEKIIIDLLMYINAGHSSGLTMMATIMHPIIMLOAKKEORIVK
KRAQKQKLTLEKREMYLSQVIDETLRLVITFSLAEKASDVQMDGYIIPKGMKVL
TWFRNVHLDPEITPDPKFDPSPRMEGYTPKAGTFLPFGSHLCPGNDLAKLEISIFL
HHFLKXRVRSRNSGCPVMPFLPHNRPKDNCARITRMP"
BASE COUNT      513 a      326 c      389 g      512 t
ORIGIN

```

```

alignment_scores:
  quality: 436.50      length: 504
  ratio: 1.480        gaps: 13
  percent similarity: 58.532      percent identity: 27.381

```

alignment_block:

us-09-668-482-4 x AF318501 ..

Align seg 1/1 to: AF318501 from: 1 to: 1740

```

5 AlaleuleuAlaSerAlaLeuCythrPheValLeuProleuLeuLeu 21
||||| : : : : : |||
70 GGTATGACGGAAACGGGTTGATCTCATGTGCTTCCCTTGATTAATAT 119
21 e.....LeuAlaAlaLeuLeuLeuLeuLeuLeuLeuLeuLeu 30
: : : : : |||
120 GGGATTGTTGTTGAATGGTGGTCTGAAGAGATGATGTTTGATTT 169
30 eUTyCySValSerGlyArgAspArgSerCySAlaLeuProLeuPro 46
||||| : : : : : |||
170 ATGTGTCCAACTTGTTGAGAAAAAACAATA.....CTGCACCA 210
47 GlyThrMetGlyPheProPhePheGlyGluThrLeuGluMetVal..... 61
||| : : : : : |||
211 GGTGATTGGGATGGCAGTCATAGCAACATGTGGTCTTCTTAGAGC 260
62 .....LeuGluArgArgGlySerPheLeuGluMetCysArgGlySer 75
: : : : : |||
261 TTTCAAACATCTGATCCGATCCCTTCATCCAAATCCATACACACAGGT 310
75 YrGly.....PheIleTyThrHisLeuPheGlyArgProThrVal 89
||||| : : : : : |||
311 ATGGCGGACCGGATTTATAAGCACACATGTTGGTAGCCCATGTGTA 360
90 ArgValMetGlyAlaAspAsnValArgArgIleLeuLeuGlyAspAsp 106
||| : : : : : |||
361 CTAGTAACAACACACAGACTGTAGCGAGTTTAAACAGATGATGATGAT 410
106 GluValSerValHisThrProAlaSerValArgThrIleLeuGlySerG 123
: : : : : |||
411 CTC....CACATAGGTGGCCAAATCTACATGAACCTCATTTGCAGCA 457
123 IyCySLeuSerAsnLeuHisAspSerSerHisIySGLNArgIySVal 139
: : : : : |||
458 AGTCCTTGTGTATCTCTTGAAGAACAACAGGCTCAGGCGTTTG 507
140 IleMetArgAlaPheSer...ArgIleAlaLeuGluCySThrValPro 155
: : : : : |||

```

```

508 ACTTTCCTCTCTGTCAATGGCCCGAAGCTCTCTGTATATACATACGTT 557
155 ILeThrGluGluValGlySerSerLeuGluGluIntIyPLeuSerCyS 172
||| : : : : : |||
558 CATTAAGAAACTGTTAATACATGATCTAGAAAATGGTCCAAATGGGAG 607
172 IuArgGlyLeuLeuValTyThrProGluValIySArgLeuMetPheArg 188
||| : : : : : |||
608 AATTCAGATCTTATCTCAT.....TTGGGTAGGCTTACGTTTAAGTGC 651
189 AlMetArgIleLeuLeuGlyCySGLuProGluLeuAlaGlyAspGly 205
||| : : : : : |||
652 ATATGTACATATTTCTCAGACAGTACAGAGACACATCTCATGAT..... 656
205 pSerGluGluGluLeuValGluAlaPheGluGluMetThrArgAsnLeu 222
: : : : : |||
697 .....TCATTTGACGGAGGATATACCAACCTTAACATAGGAGTTC 736
222 heSerLeuProIleAspValProPheSerGlyLeuIyArgIyGlyMet 238
: : : : : |||
737 GAGCAATGGGATTAATCTTCTGCGGTTGCTTATCATAGAGCTTTGAG 786
239 AlaArgAsnLeuIleHisAlaArgIleGluGluAsnIleArgAlaIy 255
||||| : : : : : |||
787 GCAAGGAAAAAACTGTAGCTGCCTT...CAGTCCATAGTACTTAACCG 833
255 ecySGLyLeuArgAlaSerGluAlaGlyGluGlyCySAspAlaLeuG 272
: : : : : |||
834 AACAAATCAAAAGAACAGACAGATTTTCATCCAAATGAGAAAGATATG 883
272 ILeuLeuIleGluHisSerThrPLeuArgGlyIyGluArgLeuAspMet 288
: : : : : |||
884 ATATCTAATAGATGTTAAGATGAAAGGAAAGAGCTTACATGATGAG 933
289 AlAlaLeuGluSerSerThrGluLeuLeuPheGlyIyHisGluThr 305
: : : : : |||
934 AAAATATGACCTTTGTTGATGATCTTAATGCGGATGAGAAATCTTC 983
305 rAlaSerAlaAlaThrSerLeuIleThrTyLeuGlyLeuTyProHis 322
: : : : : |||
984 TGGACACCTCAGATGTGGGCTACCATTTTGATGCAAGAACACCTATGA 1033
322 aLeuGluIySValArgGluGluLeuLeuSerIySGLyLeuLeuCyS 338
: : : : : |||
1034 TTCTGCAAAAGCTAAGAAAGAA.....CAAGAGAGATCGTAAAAAAG 1077
339 SerAsnGluAspAsnIySLeuAspMetGluIleLeuGluGluLeu 355
: : : : : |||
1078 AGAGCACCGGACAGAAAGTTGACGCTTAAGACACACGTGAATGTGTA 1127
355 rIleGlyCySValIleYsGluThrLeuArgLeuAsnProProValPro 372
: : : : : |||
1128 TTTTCTCAGGTTATGATGACACCTTCGAGTAATTCATCTCTCTGA 1177
372 IyGlyPheArgValAlaLeuIySThrPheGluLeuAsnIyGluIle 388
: : : : : |||
1178 CGGCTTTCAGGAGACAAAGAGTATGCCAATGATGATGATTAAT 1227
389 ProYsGlyTyPAsnValIleTySerIleCySAspThrHisAspVal 405
||||| : : : : : |||
1228 CCAAAAGGCTGGAAGTCTGACTGTTAGGAAGCTCATTTGGACCC 1277
405 aGluIlePheThrAsnIySGLuGluPheAsnProAspArgPheSerAla 422
||||| : : : : : |||
1278 TGAATCTACCCGATCCAAAGAAATTTGATCTTCATAGTGGAGGAT 1327
422 rHisProGluAspAlaSerArgPheSerPheIleProPheGlyIyGly 438
||||| : : : : : |||
1328 ACACACCAAAACAGGC.....ACATTCCTTCCTTTTGTTGGGA 1368
439 IeuArgSerCySValGlySGLuPheAlaIySLeuLeuIySLeuIle 455
: : : : : |||
1369 AGCCACCTATGCCCGGAAACGATCTTGCAAGGCTGAGATTTCCATTTT 1418

```

```
455 ethrvalgluleualarghiscysasptprglnleuasn...GlyP 471
|      :::::      :::      |||      |||
1419 TCTTCATCATTTCTCTCCTCAATACCGGGTGGAAAGAGCAATCCCGAT 1468
471 roprothrmellythrserprothrvaltyrprovalaspasnleupro 487
|      :::::      |||      |||      |||
1469 GTCCGGGTGATG...TTCTTGCTTCACCAATCGACCCAAAGATTAATTGCTTA 1515
488 AlargPheThr 491
|      :::::      |||
1516 GCAAGATCACC 1527
```


Align seg 1/1 to: AAV09247 from: 1 to: 1494

```

1 MetGlyLeuProAlaLeuLeuAlaSerAlaLeuCysThrPheValLeuPr 17
  |||
1 ATGGGGCTCCGGCGCTGGCCAGTGGCTCTCCACCTTGCTGCTGCG 50
  |||
17 OleuLeuPheLeuAlaAlaIleLysLeuTrpAspLeuTyrCysValS 34
  |||
51 GCTGCTGCTCTTCTGCTGGCATCAAGCTCGGACCTGCTACTCGTGA 100
  |||
34 ergLyArgAspArgSerCysAlaLeuProLeuProProGlyThrMetGly 50
  |||
101 GCGGCGGAGCCGAGTGTGGCTCCCATTTGCCCCCGGAGCTATGGGC 150
  |||
51 PheProPhePheGlyLysThrLeuGlnMetValLeuGlnArgArgLysph 67
  |||
151 TTCCCTCTTTGGGAAACCTTGACATGGTACTGACGCGAGGAAAGTT 200
  |||
67 eleuGlnMetLysArgArgLysTyrGlyPheIleTyrLysThrHisLeuP 84
  |||
84 heGlyArgProThrValArgValMetGlyAlaAspAsnValArgArgIle 100
  |||
251 TGGGGCGCCACCGTACGGGGTGGTGGGCGGACAAATGTGGCGCATC 300
  |||
101 LeuLeuGlyAspAspArgLeuValSerValHisTrpProAlaSerValAr 117
  |||
301 TTGCTCGGAGAGACCGGCTGGTGTCCGCTCCACGCGGCGCTGGTGTG 350
  |||
117 gThrIleLeuGlySerGlyCysLeuSerAsnLeuHisAspSerSerHisL 134
  |||
351 CACCATTTGGGATCTGGCTGCTCTTAACCTGCACGACTCTTGACACA 400
  |||
134 YSGlnArgLysLysValIleMetArgAlaPheSerArgLysAlaLeuGln 150
  |||
401 AGCAGCGCAAGAGGTGATTTATGGGGCTTCAGCCGCGAGCATTCGAA 450
  |||
151 CysTyrValProValIleThrGlnGluValGlySerSerLeuGlnTr 167
  |||
451 TGCCTACGTGCGGTGATCACCGAGAAAGTGGCGACAGCTTGAGACAGT 500
  |||
167 PleuSerCysGlyGluArgGlyLeuLeuValTyrProGluValLysArgL 184
  |||
501 GCTGAGCTGGCGAGCGGGCTCTGCTGCTACCCCGAGGTGAAGCGCG 550
  |||
184 eumetPheArgIleAlaMetArgIleLeuLeuGlyCysGlnProGlnLeu 200
  |||
551 TCATGTCTCGAATGCGCATGCGCATCTACTGGGCTGGCAACCCACAGT 600
  |||
201 AlaGlyAspGlyAspSerGlnGlnGlnLeuValGluAlaPheGlnGlu 217
  |||
601 GCGGCGGAGGGGAGCTCCGACACAGAGCTTGTGAGAGGCGCTTCGAGAAAT 650
  |||
217 tThrArgAsnLeuPheSerLeuProIleAspValProPheSerGlyLeuT 234
  |||
651 GACCGCGCAATCTTCTGCTGCTGCCATGACGTGCCCTTCAGCGGGCTGT 700
  |||
234 yrrArgGlyMetLysAlaArgAsnLeuIleHisAlaArgIleGlnGlnAsn 250
  |||
701 ACCGGGGCATGAGGCGGGAACCTCATTTACGCGCGCATCGAGAGAAC 750
  |||
251 IleArgAlaLysIleCysGlyLeuArgAlaSerGluAlaGlyGlnGlyC 267
  |||
751 ATTCGCGGCAAGATCTGGCGCTGGGCGCATCCGAGGGGCGGAGGGCTG 800
  |||
267 sLysAspAlaLeuGlnLeuLeuIleGluHisSerTrpGluArgGlyGlu 284
  |||
801 CAAGACCGCGCTGCACTGTTGATGACCACTCGTGGAGAGAGGAGAGAC 850
  |||
284 rGleuAspMetGlnAlaLeuLysGlnSerSerThrGluLeuLeuPheGly 300
  |||
851 GGTGTGACATGAGGCACTAAAGCAATCTTCAACGAACTCCTCTTGG 900
  |||

```

```

301 GLYHISGLUThrThrAlaSerAlaAlaThrSerLeuIleThrTyrLeuG1 317
  |||
901 GGACACGAAACCAACGCGCCAGTGCACACCATCTCTGATCATTACTCTGG 950
  |||
317 yLeuTyrProHisValLeuGlnLysValArgGluLeuLysSerLysG 334
  |||
951 GCTTACCACATGTTTCCAGAAAGTGGAGAGAGAGGAGAGAGTAAGG 1000
  |||
334 yLeuLeuCysLysSerAsnGlnAspAsnLysLeuAspMetGluLeu 350
  |||
1001 GTTTACTTTGCAAGAGCAATCAAGACAAAGTTGAGACTGGAAATTTTG 1050
  |||
351 GluGlnLeuLysTyrIleGlyCysValIleLysGluThrLeuArgLeuS 367
  |||
1051 GAACAACCTTAATACATCGGGGTGTATTATTAGAGACCTTCGACTGAA 1100
  |||
367 nProProValProGlyGlyPheArgValAlaLeuLysThrPheGluLeu 384
  |||
1101 TCCCCAGTTCCAGAGAGGTTTCGGGTGCTCTGAAAGACTTTTGAAATPA 1150
  |||
384 sngLYTyrGlnIleProLysGlyTrpAsnValIleTyrSerIleCysAsp 400
  |||
1151 ATGATACCAAGATTTCCCAAGGCTGGAATGTTATCTACATATCTGTGAT 1200
  |||
401 ThrHisAspValAlaGluIlePheThrAsnLysGluGluPheAsnProAs 417
  |||
1201 ACTCATGATGTGCGAGAGATCTTACCAACAAGAAATTTAAATCTGTA 1250
  |||
417 pArgPheSerAlaProHisProGluAspAlaSerArgPheSerPheIle 434
  |||
1251 CCGATTCAAGTGTCTCTCACCCAGAGATGATCCAGGTTCACCTTCATTC 1300
  |||
434 roPheGlyGlyLeuArgSerCysValGlyLysGluPheAlaLysIle 450
  |||
1301 CATTTGAGAGAGCGCTTAGAGCTGTAGAGCAAGAATTTCCAAAAAT 1350
  |||
451 LeuLeuLysIlePheThrValGluLeuAlaArgHisCysAspTrpGlnLe 467
  |||
1351 CTTTCAAAATATTTACAGTGGAGCTGGCGACATTTGAGCTGCGAGCT 1400
  |||
467 uLeuAsnGlyProProThrMetLysThrSerProThrValTyrProVal 484
  |||
1401 TCTTAATGAGACTCTTACATGAAAAACAGTCCACCGTGATCTCTGTGG 1450
  |||
484 sPAsnLeuProAlaArgPheThrHisPheHisGlyGluIle 497
  |||
1451 ACAATCTCCCTCAGAGATTCACCCATTTCCATGAGGGAATTC 1491
  |||
seq_name: /SIDSL/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV12204
seq_documentation_block:
ID AAV12204 standard; cDNA; 1494 BP.
XX
AC AAV12204:
XX
DT 22-JUN-1998 (first entry)
XX
DE Human retinoid metabolizing protein hp450RA1 cDNA.
XX
KW Retinoid metabolizing protein; P450RA1; retinoid oxidase;
KW Retinoic acid; human; inhibitor; antisense; cancer;
KW actinic keratosis; oral leukoplakia; head tumour; neck tumour;
KW non-small cell lung carcinoma; basal cell carcinoma;
KW acute promyelocytic leukaemia; skin cancer; acne; psoriasis;
KW Ichthyosis; therapy; diagnosis; screening; cytochrome P450; ss.
XX
OS Homo sapiens.
XX
PN W09749815-A1.
XX
PD 31-DEC-1997.
XX

```

```
PF 23-JUN-1997: 97MO-CA00440.
XX
PR 01-OCT-1996: 96US-0724466.
PR 21-JUN-1996: 96US-0667546.
XX
PA (TOOH ) UNIV QUEENS KINGSTON.
PI Beckett BR, Jones G, Petkovich PM, White JA;
DR MPI: 1998-071178/07.
XX P-PSDB: AAM44160.
XX Retinoid metabolising protein - useful to develop products to treat,
PT e.g. cancer, actinic keratosis, oral leukoplakia, acne, psoriasis or
PT ichthyosis
XX
PS Claim 15; Page 55-57; 110pp; English.
XX
XX This cDNA clone codes for a novel human retinoid metabolising
CC protein (see AAM44160) designated hp450RAI. The encoded protein is a
CC retinoid oxidase that has the ability to hydroxylate retinoic acid
CC at the 4 position of the beta-ionone ring and is inducible in
CC epithelial cells exposed to retinoic acid. To isolate the clone,
CC zebrafish p450RAI was used to search an expressed sequence tag
CC database, and an isolated clone was used to screen a cDNA library
CC generated from an M2 cell line treated with retinoic acid. The
CC hp450RAI gene has been localised to 10q23-24. Zebrafish, human
CC and mouse p450RAI cDNA clones (see AAV12203-05) are claimed. They
CC can be used in the recombinant production of p450RAI. Antisense
CC nucleic acids can be used in a claimed method for inhibiting
CC retinoic acid hydroxylation for the treatment of cancer, actinic
CC keratosis, oral leukoplakia, secondary tumours of the head and/or
CC neck, non-small cell lung carcinomas, basal cell carcinomas, acute
CC promyelocytic leukaemia, skin cancer, and premalignancy associated
CC with actinic keratosis, acne, psoriasis and/or ichthyosis.
CC Promoter sequences (see AAV12206-08) for the p450RAI genes are also
CC claimed.
XX
XX Sequence 1494 BP; 332 A; 414 C; 430 G; 318 T; 0 other:
SQ
alignment_scores:
Quality: 2611.00 Length: 497
Ratio: 5.254 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-668-482-4 x AAV12204 ..
Align seg 1/1 to: AAV12204 from: 1 to: 1494
1 MetGlyLeuProAlaLeuLeuAlaSerAlaLeuCysThrPheValLeuPr 17
|||||
1 ATGGGGCTCCCGGGGCTGCTGGCCAGTGGCTGTGCACCTTCGTGCTGCC 50
17 cLeuLeuLeuPheLeuAlaAlaIleuLeuLeuTPASPLeuTyrCysVal 34
|||||
51 GCTGCTCTCTTCTCTGCTCGATCAAGCTCTGGACCTGTACTGCTGA 100
34 erGlyAArgSPArgSerCysAlaLeuProLeuProGlyThrMetGly 50
|||||
101 GCGGCGCGGACGCGAGTGTGCTCCCTCCATGCCCCCGGAGACTATGGC 150
51 PheProPhePheGlyGluThrLeuGluMetValLeuGluAlaArgGlySph 67
|||||
151 TTCCCTCTCTTTGGGAAACCTTGCAATGTGTACTGACGCGAGGAAGTT 200
67 eLeuGluMetLeuArgArgGlySphGlyPheIleTyrThrHisLeuP 84
|||||
201 CCTGCAGTGAAGCGCAGGAATACGCTTCATCTACAAAGCGCATCTGT 250
84 heGlyArgProThrValArgValMetGlyAlaAspAsnValArgArgIle 100
|||||
```

```
251 TCGGCGCGCCACCGTACGGGTGATGGCGCGGACAAATGTGGCGGCATC 300
101 LeuLeuGlyASPArgLeuValSerValHisTrpProAlaSerValAr 117
|||||
301 TTGCTCGAGACGACCGGCTGTGTCGCTCCACTGCGCCACGTCGGGCGC 350
117 gThrIleLeuGlySerGlyCysLeuSerAsnLeuHisASPserSerHisL 134
|||||
351 CACCAATCTGGATCTGGCTGCTCTTAACCTGCACGACTCTCTGCACA 400
134 ySGlnArgLySLysValIleMetArgAlaPheSerArgGluAlaLeuGlu 150
|||||
401 AGCAGCGCAAGAAAGTGATTTATGGGGCTTCACCGCGAGGCTGAA 450
151 CysTyrValProValIleThrGluGluValGlySerSerLeuGluGlnTr 167
|||||
451 TGCTACGTGCGCGGTGATCACCGAGGAAGTGGCGAGCGCTGGAGCAGTG 500
167 pLeuSerCysGlyGluArgGlyLeuLeuValTyrProGlyValLysArgL 184
|||||
501 GCTGAGCTGGCGGCGAGCGCGCTCTCTGTGTTACCCGAGGTGAAGCGCC 550
184 eUmeTPheArgIleAlaMetArgIleLeuLeuGlyCysGluProGluLeu 200
|||||
551 TCATGTTCCGAATCGCCCATGGCCATCTTACTGGCTCGGAACCCCACTG 600
201 AlaGlyASPArgLysSerGlyGluGlnLeuValGluAlaPheGluGluMe 217
|||||
601 GCGGCGAGGGGAGCTCCGACGACGAGCTTGTGAGGCTTCGAGGAAT 650
217 ThrArgAsnLeuPheSerLeuProIleASPValProPheSerGlyLeuT 234
|||||
651 GACCCGCAATCTCTTCTGCTGCCATGCCATGACGAGTGCCTTCAGCGGGCTG 700
234 YrArgGlyMetLysAlaArgAsnLeuIleHisAlaArgIleGluGlnAsn 250
|||||
701 ACCGGGGCATGAAGCGCGGAACCTCATTCACGCCGCGATGAGAGCAAC 750
251 IleArgAlaLysIleCysGlyLeuArgAlaSerGluAlaGlyGlnGlyCy 267
|||||
751 ATTGCGGCCAAGATCTCGGGCTGGCGGATCCGAGGGGCGGCGAGGCTG 800
267 sLysASPAlaLeuGlnLeuLeuIleGluHisSerTrpGluArgGlyGluA 284
|||||
801 CAAAGCGGCGTGCAGCTGTGATGACGACTCGTGGAGAGAGGAGAGC 850
284 rGluASPMetGlnAlaLeuLysGlnSerSerThrGluLeuLeuPheGly 300
|||||
851 GGCTGGACATGCGAGCGCACTTAAGCAATCTTCAACCGAATCTCTTTTGA 900
301 GlyHisGluThrThrAlaSerAlaIleThrSerLeuIleThrTyrLeuG 317
|||||
901 GGACACGAAACCAAGCGCCAGTGCAGCCACATCTCTGATCATTACTTGGG 950
317 yLeuTyrProHisValLeuGlnLysValArgIleGluLeuLysSerLysG 334
|||||
951 GCTCTACCCACATGTTCTCCAGAAAGTGCAGAGAAGAGCTGAAGATAAGG 1000
334 lYLeuLeuCysLysSerAsnGlnAspAsnLysLeuAspMetIleuLeuLeu 350
|||||
1001 GTTTACTTTGCCAAGAACCAATCAAGACAAAGTGGACATGGAAATTTTG 1050
351 GluGlnLeuLysTyrIleGlyCysValIleLysGluThrLeuArgLeuAs 367
|||||
1051 GAACAACTTAATACATCGGGTGTCTTATTAAAGAGACCCCTTCGACTGAA 1100
367 nProProValProGlyGlyPheArgValAlaLeuLysThrPheGluLeuA 384
|||||
1101 TCCCCAGTTCACGAGAGGTTTCGGGTCTGCTCGAAGACTTTTGAATTAA 1150
384 snGlyTyrGlnIleProLysGlyTyrPheValIleTyrSerIleCysAsp 400
|||||
1151 ATGGATACCAAGATTCCCAAGGCTGGAATGTATTCTACAGTATCTGTGAT 1200
```

```
401 ThrHisAspValAlaGluIlePheThrAsnLysGluGluPheAsnProAs 417
|||||
1201 ACTCATGATGTGGCAGAGATCTTCCACCAAGAAAGATTAACTCTGA 1250
417 PArgPheSerAlaProHisProGluAspAlaSerArgPheSerPheIleP 434
|||||
1251 CCGATTTCAGTGCCTCCTCACCCAGAGATGCATCCAGGTTCACCTCATTC 1300
434 rOPheGlyGlyGlyLeuArgSerCysValGlyGlyGluPheAlaLysIle 450
|||||
1301 CATTTGGAGAGGACCTTAGAGACTGTGAGGCCAAAGAAATTCAAAAAT 1350
451 LeuLeuLysIlePheThrValGluLeuAlaArgHisCysAspTrpGlnLe 467
|||||
1351 CTTCCTCAAAATATTTCACAGTGGAGCTGGCCAGGCATGTGACTGGCAGCT 1400
467 uLeuAsnGlyProProThrMetLysThrSerProThrValIlyProValA 484
|||||
1401 TCTAATATGACCTCCTACATGAAACCAAGTCCACCGGTATCCTGTGG 1450
484 sPAsnLeuProAlaArgPheThrHisPheHisGlyGluIle 497
|||||
1451 ACAATCTCCCTCCAGATTCACCCCATTTCCATGGGGAATC 1491
seq_name: /SIDSL/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA60752
seq_documentation_block:
ID AAA60752 standard; cDNA; 2124 BP.
XX
AC AAA60752:
XX
DT 31-OCT-2000 (first entry)
XX
DE Human PSEC64 encoding nucleotide sequence SEQ ID NO:1.
XX
KW Human; PSEC64; neuron growth; nerve disease; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 46..696
FT /*tag= a
FT /product= "PSEC64"
FT /note= "protein related to neuron growth"
XX
JP2000152790-A.
XX
PN 06-JUN-2000.
XX
PD 19-NOV-1998; 98JP-0329989.
XX
PF 19-NOV-1998; 98JP-0329989.
XX
PR 19-NOV-1998; 98JP-0329989.
XX
PA (HERI-) HERIKKUSU KENKYUSHO KK.
XX
DR WPI: 2000-468126/41.
XX
P-PSDB: AAB12489.
XX
PT A protein related to the growth of neuron and a gene encoding said
XX protein -
XX
PS Claim 2; Page 8-10; 13pp; Japanese.
XX
CC The present sequence encodes a human protein, designated PSEC64, which
XX is related to neuron growth. The PSEC64 protein and its gene can be used
XX for the development of a preventive agent for use in the treatment of
XX diseases in which nerves are involved.
SQ Sequence 2124 BP; 509 A; 510 C; 547 G; 558 T; 0 other;
Alignment_scores:
```

```
Quality: 2562.00 Length: 497
Ratio: 5.186 Gaps: 0
Percent Similarity: 99.396 Percent Identity: 99.195
alignment_block:
US-09-668-482-4 x AAA60752 ..
Align seg 1/1 to: AAA60752 from: 1 to: 2124
1 MetGlyLeuProAlaLeuAlaSerAlaLeuCysThrPheValLeuPr 17
|||||
46 ATGGGCTCCCGCGCGTGGCCAGTGGCCCTGCACCTTCGACCTTGCTGGC 95
17 oLeuLeuPheLeuAlaAlaIleLysLeuTrpAspLeuTrpCysValS 34
|||||
96 GCTGCTGCTCTTCCTGCTGCTGCATCAAGCTCTGGACCTGTACTGGCTA 145
34 eGlyArgAspArgSerCysAlaLeuProLeuProGlyThrMetGly 50
|||||
146 GCGGCGGCGACCGCAGATTGTGCCCTCCATTCGCCCGCGGACTATGGGC 195
51 PheProPhePheGlyGluThrLeuGlnMetValLeuGlnArgArgLysP 67
|||||
196 TTCCCTCTCTTGGGAAACCTTGCAGATGTACTGCACGAGGAAAGTT 245
67 eLeuGlnMetLysArgArgLysTyrglyPheIleTyrlsThrHisLeuP 84
|||||
246 CCGCAGATGAAAGCCAGGAATACGGCTTCATCAAGACGACATCTGT 295
84 heGlyArgProThrValArgValMetGlyAlaAspAsnValArgArgIle 100
|||||
296 TCGGGCGGCGCCACCGTACGGGTGATGGGCGGACAAATGTGGCGGCAATC 345
101 LeuLeuGlyAspAspArgLeuValSerValHisTrpProAlaSerValAr 117
|||||
346 TTGCTCGGAGACACCGGCTGCTGCTGCCCTGCCACCGCAGCTCGGTGC 395
117 gThrIleLeuGlySerGlyCysLeuSerAsnLeuHisAspSerSerHisL 134
|||||
396 CACCATTTCTGGGATCTGGCTGCTCTTAACCTGCACGACCTCTCCGCACA 445
134 ysgLnaArgLysLysValIleMetArgAlaPheSerArgGluAlaLeuGlu 150
|||||
446 AGCAGCGAAGAGAGGTGATTAAGCGGCTTCAAGCGGAGGACCTCGAA 495
151 CysTyrrValProValIleThrGluGluValGlySerSerLeuGluGlnTr 167
|||||
496 TGCTACGTGCCGGTGATCACCGAGAGAGTGGGCAAGCCTTGAGACAGTG 545
167 PleuSerCysGlyGluArgGlyLeuLeuValIlyTrpGluValLysArgL 184
|||||
546 GCTGAGCTGGCGCGAGCGCGGCTCTGCTTACCCCGAGGTGAAGCCGC 595
184 eumetPheArgIleAlaMetArgIleLeuLeuGlyCysGluProGlnLeu 200
|||||
596 TCATGTTCCGAATCGCCATGCGCATCTACTGGGCTGGGAACCCCAATG 645
201 AlaGlyAspGlyAspSerGluGlnGlnLeuValGluAlaPheGluGlu 217
|||||
646 GCGGCGCAGCGGGAAGCTCCGAGCAGACCTTGTGAGGCTTCGAGGA. AT 694
217 tThrArgAsnLeuPheSerLeuProIleAspValProPheSerGlyLeuT 234
|||||
695 GACCCGCAATCTCTTCTGCTGCCATGACGTGCTTCAGCGGGGCTGT 744
234 yTrArgGlyMetLysAlaArgAsnLeuIleHisAlaArgIleGluGlnAsn 250
|||||
745 ACGGGGCGATGAAGCGCGGAACCTCATTCACGCGCGCATGAGAGAAC 794
251 IleArgAlaLysIleCysGlyLeuArgAlaSerGluAlaGlyGlnGlyCy 267
|||||
795 ATTGCGCGCAAGATCTGCGGGCTGCGGCGATCCGAGCGGCGGCGGCTG 844
```


[illegible]

FH	Key	Location/Qualifiers
FT	CDS	24..1518 +/tag= ^a /product="Cytochrome P450RAI isoform"
FT		
FT		
XX		
XX	M09749833-AZ.	
PX		
PD	31-DEC-1997.	
XX		
XX	23-JUN-1997:	97MO-CA00488.
XX		
PR	01-OCT-1996;	96US-0724466.
PR	21-JUN-1996;	96US-0667546.
XX	(TOOH) UNITV QUEENS KINGSTON.	
PA		
XX		
PI	Pelkovich PM.	
XX		
DR	WPI: 1998-077193/07.	
DR	P-PADB; AAM37735.	
PT		
PT	Identifying DNA encoding inducible or suppressible cytochrome P450 -	
PT	by screening for drugs which reduce the catabolism of retinoic acid,	
PT	useful in cancer chemotherapy and the treatment of acne and	
PS	psoriasis	
XX		
FS	Disclosure; Pages 59D-59F; 113pp; English.	
XX		
CC	This nucleotide sequence encoding an isoform of cytochrome P450RAI.	
CC	Its expression is dependent on the presence of retinoic acid (RA).	
CC	The retinoid-regulated genes such as the inducible cytochrome P450RAI	
CC	gene specifically metabolises a derivative of the RA. The cytochrome	
CC	P450 gene in general produces enzymes involved in the oxidative	
CC	metabolism of endogenous and exogenous compounds. The cytochrome P450	
CC	nucleotide sequence can be used to induce or suppress the expression	
CC	of its protein. P450RAI is highly induced by RA in cell lines and	
CC	tissues. This allows for development of a drug screen using promoters	
CC	and nucleotide sequences to identify drugs which are useful for	
CC	reducing the catabolism of RA.	
XX		
SQ	Sequence 1725 BP; 408 A; 425 C; 475 G; 417 T; 0 other;	
alignment_scores:		
	Quality: 2437.00	Length: 497
	Ratio: 5.035	Gaps: 0
	Percent Similarity: 97.384	Percent Identity: 93.360
alignment_block:		
US-09-668-482-4 x AAV09252 ..		
Align seg 1/1 to: AAV09252 from: 1 to: 1725		
1	MetGlyLeuProAlaLeuLeuAlaSerAlaLeucyThrPheValLeuPr 17	
25	ATGGAGCGTCCCGGCCTGTGGCAGTGACGTTCGACCTTCGTGCC 74	
17	oleuleuLeuPheleuAlaAlaLilelysleuTrpaspLeuTrcyvalS 34	
75	GTGCGTGCCTTTCTCTGGCGGCGCTCAAAGCTCTGGAGACTGTACTGTGA 124	
34	erGlytArgAsparSerCysAlaLeuProLeuProGlylThrMetGly 50	
125	GCAGCGCGATCGACACTGCGCCCTTCCCTCCCCCGGATCATCATGGGC 174	
51	PheProPhePheligluThrleuglnmetVallleuglnargAglgylsph 67	
175	TTCOCATTCTTGTGGGAAACAATTGCAGATGGGGCTTCAGCGGAGAAGTT 224	
67	eLeuGlmetLevArGArgLystrigLYpheLIetYrlYstThrlslEup 84	
225	TCGTGAGATGAAAGCGCAGGAATAAGCGCTTCACTCACAAAGACGATCGT 274	

```

84 heglYArgProthrValAlaValMetGlyAlaAspAsnValArgArgIle 100
|||||
275 TTGGGGCGGCCACGGTGGCGGTGATGGCGCGGATATATGCGGGCCATC 324
|||||
101 LeuLeuGlyAspAspArgLeuValSerValHisTTPProAlaSerValAr 117
|||||
325 TTGGTGGGAGACCGCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 374
|||||
117 gThrIleLeuGlySerGlyCysLeuSerAsnLeuHisAspSerSerHisL 134
|||||
375 CACATCCCGGGCGCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 424
|||||
134 ySglArGlyLysValIleMetArGAlaPheSerArgIuAlaLeuGlu 150
|||||
425 AGCAGCGAAGAGAGGTATATGACGAGCGCTTCAGCCGCGAGCACCTCAG 474
|||||
151 CysTyValAlProValIleThrGluGluValGlySerSerLeuGluGlnTr 167
|||||
475 TGGTACGTGCTGTGATCGCTGAGAGAGACAGCATGTGCTGGAGCAGTg 524
|||||
167 pLeuSerCysGlyGluArgGlyLeuLeuValTyProGluValLysArgL 184
|||||
525 GCTAGCTGGCGCGAGCGCGCGCTCTGGTACCAGCGAGTGAAGCGCG 574
|||||
184 eumEPhaArgIleAlaMetArgIleLeuLeuGlyCysGluProGluLeu 200
|||||
575 TCATGTCCGATCCGATCGCATGCGCATCTGCTGGCGTGGAGCGGAGTCCA 624
|||||
201 AlaGlyAspGlyAspSerGluGlnGlnLeuValGluAlaPheGluGluMe 217
|||||
625 GCGGGCGGGCGGAGAGAGAGACAGCTCGTGGAGGCTTTGAGAGAGAT 674
|||||
217 tThrArgAsnLeuPheSerLeuProIleAspValProPheSerGlyLeuT 234
|||||
675 GACCCGCAATCTCTCTCTCTCCATTCAGCTGGCTTACGGCGCTGT 724
|||||
224 YrArgGlyMetLysAlaArgAsnLeuIleHisAlaArgIleGluGlnAsn 250
|||||
725 ACCGGGGCGTGAAGCGCGAACCCTTATACGCGCGCATCGAGAGAAC 774
|||||
251 IleArgAlaLysIleCysGlyLeuArgAlaSerGluAlaGlyGluGlyCY 267
|||||
775 ATTGGCGCCAGATCCGGCGCTTACGCTACAGCGCGGATGGGGGTGG 824
|||||
267 sLyAspAlaLeuGlnLeuLeuIleGluHisSerTProIuArgGlyGluA 284
|||||
825 CAAGGACGCGCTGCAGCTCTGATTTGACACTCGTGGAGAGGGGAGAGA 874
|||||
284 rLeuAspMetGlnAlaLeuLysGlnSerSerThrGluLeuLeuPheGly 300
|||||
875 GCGTGGATATCGAGGCACTAAACATCTCAACAGAGCTCTCTTTTGGT 924
|||||
301 GlyHisGluThrThrAlaSerAlaAlaThrSerLeuIleThrTyLeuG 317
|||||
925 GGTATGTAACACTACAGCCAGTGGTGGCATACATCACTTACCTAGG 974
|||||
317 YLeuTyProHisValLeuGlnLysValArgGluGlnLeuLysSerLysG 334
|||||
975 ACTTACCCACATGTCTCCAGAAAGTTCCGAAAGAGATTAAGACCAAGG 1024
|||||
334 LyLeuLeuCysLysSerAsnGlnAspAsnLysLeuAspMetGluIleLeu 350
|||||
1025 GCTTACTTTTGAAGAGCAATCAAGACAAGATTAGACATGGAACCTTGG 1074
|||||
351 GluGlnLeuLysTyrlleGlyCysValIleLysGluThrLeuAlaGluAs 367
|||||
1075 GAAGAAGCTTAATACATTTGGTGTGCTATTAAAGAGACCCGCAATTTGA 1124
|||||
367 nProProValProGlyGlyPheArgValAlaLeuLysThrPheGluLeuA 384
|||||
1125 TCCTCGCGGTTCCAGAGAGGTTTGGGTTGCTCTCAAGACTTTTGAAGCTGA 1174
|||||
384 snGlyTyrlleProLysGlyTyrlleProLysValIleTyrlleSerIleCysAsp 400

```

```

|||||
1175 ATGGATACCAAGATCCCAAGGGGCTGGATGTTATTACAGTATCTGTGAC 1224
|||||
401 ThrHisAspValAlaGluIlePheThrAsnLysGluGluPheAsnProAs 417
|||||
1225 ACCCAGATGTGGCAGATATCTTCACTTAACAAGAGAAATTAATCCGA 1274
|||||
417 pArgPheSerAlaProHisProGluLysAspAlaSerArgPheSerPheIleP 434
|||||
1275 CCGCTTATAGTCCCATCCAGAGGATGCTTCCCGGTTCAAGCTTATATTC 1324
|||||
434 ropPheGlyGlyLeuArgSerCysValGlyLysGluPheAlaLysIle 450
|||||
1325 CATTTGGAGAGAGCCCTTCGAGCTGTGTAGCCAAAGAGTTTGCAAAATTT 1374
|||||
451 LeuLeuLysIlePheThrValGluLeuAlaArgHisCysAspTrpGluLe 467
|||||
1375 CTTCCTTAAGATATTACAGTGGAGCTGGCTAGGCACTGTGATTTGGCAGCT 1424
|||||
467 uLeuAsnGlyProProThrMetLysThrSerProThrValTyProValA 484
|||||
1425 TCTAATATGACCTCCATACATGACAAGCAAGCCCACTGTGATCCCTGTGG 1474
|||||
484 sPAsnLeuProAlaArgPheThrHisPheHisGlyGluIle 497
|||||
1475 ACAATCTCCCTCAGAGATTCACCTTCTCCAGAGGAGATATTC 1515
|||||
seq_name: /SIDS1/gcdata/geneseq/geneseqn/NAL198.DAT:AAV12205
seq_documentation_block:
ID AAV12205 standard; cDNA: 1725 BP.
XX
XX AAV12205:
AC
XX
XX
DT 22-JUN-1998 (first entry)
XX
DE Mouse retinoid metabolising protein mP450RAI cDNA.
XX
XX Retinoid metabolising protein; P450RAI; retinoid oxidase;
KW retinoid acid; mouse; inhibitor; antisense; cancer;
KW actinic keratosis; oral leukoplakia; head tumour; neck tumour;
KW non-small cell lung carcinoma; basal cell carcinoma;
KW acute promyelocytic leukaemia; skin cancer; acne; psoriasis;
KW Ichthyosis; therapy; diagnosis; screening; cytochrome P450; ss.
XX
XX Mus musculus.
XX
XX
FH Key Location/Qualifiers
FT CDS 25..1518
FT /tag= a
XX
XX WO9749815-A1.
XX
XX 31-DEC-1997.
XX
XX
XX 23-JUN-1997; 97WO-CA00440.
XX
XX 01-OCT-1996; 96US-0724466.
XX 21-JUN-1996; 96US-0667546.
XX
XX (TOOH ) UNIV QUEENS KINGSTON.
XX
XX
XX Beckett BR, Jones G, Petkovich PM, White JA;
XX
XX WPI; 1998-077178/07.
XX
XX P-PDB; AAW44161.
XX
XX Retinoid metabolising protein - useful to develop products to treat,
XX e.g. cancer, actinic keratosis, oral leukoplakia, acne, psoriasis or
XX ichthyosis
XX
XX Claim 15; Page 62-64; 110pp; English.
XX

```

CC This cDNA clone codes for a novel mouse retinoid metabolizing
 CC protein (see AAM44160) designated mp450RAI. The encoded protein is a
 CC retinoid oxidase that has the ability to hydroxylate retinoic acid
 CC at the 4 position of the beta-ionone ring and is inducible in
 CC epithelial cells exposed to retinoic acid. The clone was isolated
 CC from a retinoic acid-treated P19 teratocarcinoma cDNA library
 CC using human P450RAI cDNA (see AAV12204) as probe. Zebrafish, human
 CC and mouse P450RAI cDNA clones (see AAV12203-05) are claimed. They
 CC can be used in the recombinant production of P450RAI proteins.
 CC Antisense nucleic acids can be used in a claimed method for
 CC inhibiting retinoic acid hydroxylation for the treatment of cancer,
 CC actinic keratosis, oral leukoplakia, secondary tumours of the head
 CC and/or neck, non-small cell lung carcinomas, basal cell carcinomas,
 CC acute promyelocytic leukaemia, skin cancer, and premalignancy
 CC associated with actinic keratosis, acne, psoriasis and/or
 CC ichthyosis. Promoter sequences (see AAV12206-08) for the P450RAI
 CC genes are also claimed.

XX Sequence 1725 BP, 408 A; 425 C; 475 G; 417 T; 0 other:

alignment_scores:

Quality: 2437.00 Length: 497
 Ratio: 5.035 Gaps: 0
 Percent Similarity: 97.384 Percent Identity: 93.360

alignment_block:

US-09-668-482-4 x AAV12205 ..

Align seg 1/1 to: AAV12205 from: 1 to: 1725

1 MetGlyLeuProAlaLeuAlaSerAlaLeuCysThrPheValLeuPr 17
 |||||
 25 ATGGGGCTCCGGCGCTGCGCAGTGCCTCTGCACCTTCGTGCGCC 74
 |||||
 17 OleuLeuPheLeuAlaIleLeuLeuTrpAspLeuTrpCysValS 34
 |||||
 75 GCTGCTCTCTCTCTGCGCGCTCAAGCTCTGGACCTTACTGTCTGA 124
 |||||
 34 erGlyArGAspArGSerCysAlaLeuProLeuProGlyThrMetGly 50
 |||||
 125 GCACCGCGATCGACGCTGCGCCCTCCCTGCGCCCGGTACCATTGGC 174
 |||||
 51 PheProPhePheGlyGluThrLeuGluMetValLeuGluArgArgLys 67
 |||||
 175 TTCCTATTCTTTGGGGAACATTGCAGATGCTGCTTACGCGAGAGATT 224
 |||||
 67 eLeuGluMetLysArGArgLysTrpGlyPheIleTrpLysThrHisLeuP 84
 |||||
 225 TCTGCAGATGAAGCGGAGAAATACGGCTTCATCTACAGCGCATCTGT 274
 |||||
 84 heGlyArProThrValArgValMetLysAlaAspAsnValArgArgLle 100
 |||||
 275 TTGGCGCGCGCCACGCTGCGGTGATGGCGCGGATTAATGTGCGCGCATC 324
 |||||
 101 LeuLeuGlyAspAspArgLeuValSerValHisTrpProAlaSerValAr 117
 |||||
 325 TTGGCTGGAGAGACCGGTTGGTCTCGGTGACAGCGCGGCGGTGGCG 374
 |||||
 117 gThrIleLeuGlySerGlyCysLeuSerAsnLeuHisAspSerSerHisL 134
 |||||
 375 CACCATCTCTGGCGCTGCTCTCCAACTGCACGATTCCTCGCACA 424
 |||||
 134 yGluGluArgLysValIleMetArgAlaPheSerArgGluAlaLeuGlu 150
 |||||
 425 AGCACCGGAAAGAGGTATATGACGCTTCACCGCGGAGGACCTCCAG 474
 |||||
 151 CysTrpValProValIleThrGluGluValGlySerSerLeuGluGlnTr 167
 |||||
 475 TGCATCGTGTCTGATGCTGAGGAAGTCAAGCAAGTTGTCTGGAGCAAGTG 524
 |||||
 167 pleuSerCysGlyGluArgGlyLeuLeuValTyrProGluValLysArgL 184
 |||||

525 GCTAAGCTCGCGGAGCGCGGCTCTGTTCTACCCCGAGGTGAAGCGCC 574
 |||||
 184 eumETPheArgIleAlaMetArgLleLeuLeuGlyCysGluProGluLeu 200
 |||||
 575 TCATGTTCCGATCGCCATGCGCATCTCGCTGGGCTGGAGCCGGGTCCA 624
 |||||
 201 AlaGlyAspGlyAspSerGlyGluGluLeuValGluAlaPheGluGluMe 217
 |||||
 625 GCGGCGCGCGGGAGAGACAGACACACAGCTCTGGAGCGTTTCGAGGAGAT 674
 |||||
 217 tThrArgAsnLeuPheSerLeuProIleAspValProPheSerGlyLeuT 234
 |||||
 675 GACCGCATCTCTCTCTCTCCATGACGTGCGCTTATGCGCGCTGT 724
 |||||
 234 yTrArgGlyMetLysAlaArgAsnLeuIleHisAlaArgIleGluGluAsn 250
 |||||
 725 ACCGGGCGGTGAAGCGCGGAACTTATACACCGCGCATGAGAGAGAAC 774
 |||||
 251 IleArgAlaLysIleCysGlyLeuArgAlaSerGluAlaGlyGlnGlyCy 267
 |||||
 775 ATTGCGCCCAAGATCCGCGGCTTCAGGCTACAGACCGGATGGGGCTTG 824
 |||||
 267 sLysAspAlaLeuGluLeuLeuIleGluHisSerTrpGluArgGlyGluA 284
 |||||
 825 CAAGAGCGCGTGCAGCTCTGATTGACACTCTGTGGAGAGGGAGAGA 874
 |||||
 284 rGLeuAspMetGluAlaLeuLysGlnSerSerThrGluLeuLeuPheGly 300
 |||||
 875 GCGTGAATGACAGCGACTAAACATGCTCAACAGAGCTCTCTTTGTGT 924
 |||||
 301 GlyHisGluThrThrAlaSerAlaIleThrSerLeuIleThrTyrLeuGl 317
 |||||
 925 GGTCAATGAACATACAGCCAGTCTCGACATCATGATCACTTACTTACGTAG 974
 |||||
 317 yLeuTrpProHisValLeuGluLysValArgGluGluLeuLysSerLysG 334
 |||||
 975 ACTCTACCCACATGTCCTCCAGAAAGTTGAGAGAGATAAAGACCAAGG 1024
 |||||
 334 yLeuLeuLysLysSerAsnGluAspAsnLysLeuAspMetGluLeuLeu 350
 |||||
 1025 GCTTACTTTGCAGAGCAATCAAGACACAACTTGAGTGAACCTTGG 1074
 |||||
 351 GluGluLeuLysTyrIleGlyCysValIleLysGluThrLeuArgLeuAs 367
 |||||
 1075 GAACAGCTTAATAATACATGGGTGTGTCATTAAAGAGACCTTCGATTGAA 1124
 |||||
 367 nProProValProGlyGlyPheArgValAlaLeuLysThrPheGluLeuA 384
 |||||
 1125 TCTTCGGTTCAGAGGAGGTTCGGTGTCTGGAAGACTTTTGAAGCTGA 1174
 |||||
 384 snGlyTyrGluIleProLysGlyTrpAsnValIleTyrSerIleCysAsp 400
 |||||
 1175 ATGGATACCGAGATCCCAAGCGCTGGAATGTTATTACAGTATCTGTGAC 1224
 |||||
 401 ThrHisAspValAlaGluIlePheThrAsnLysGluGluPheAspProAs 417
 |||||
 1225 ACCCAGATGTGCAGATATCTTCACTAACAGAGAGAAATTAAATCCCA 1274
 |||||
 417 PArgPheSerAlaProHisProGluAspAlaSerArgPheSerPheIleP 434
 |||||
 1275 CCGCTTTATAGTCCCTCATATCCAGAGATGTTCCGCTTCAGCTTATTC 1324
 |||||
 434 ropheGlyGlyGlyLeuArgSerCysValGlyLysGluPheAlaLysIle 450
 |||||
 1325 CATTTGAGAGAGCGCTTCGAGCTGTAGGCAAAAGTTTGCAAAAAATY 1374
 |||||
 451 LeuLeuLysIlePheThrValGluLeuAlaArgHisCysAspTrpGlnLe 467
 |||||
 1375 CTTCCTTAAGATATTACAGTGAAGCTGGCTAGGCACTGTGATGGCAGCT 1424
 |||||
 467 uLeuAsnGlyProProThrMetLysThrSerProThrValTyrProValA 484
 |||||
 1425 TCTTAATGAGACTCTCATATGAAGAACCCACACTGTGTACCTGTGG 1474
 |||||

484 spasnleuproAlaArpHehrHisPheHisGlyGluIle 497
|||||
1475 ACAATCTCCCTCAGATTCACCTACTTCCAGAGGATATC 1515

seq.name: /SUSD1/gcdata/geneseq/geneseqn/NA1998.DAT:AAV09251

seq_documentation_block:

ID AAV09251 standard; cDNA; 1850 BP.

XX AAV09251;

XX 07-JUL-1998 (first entry)

XX Cytochrome zp450RAI cDNA sequence.

XX Retinoid regulated gene; cytochrome P450 gene; enzyme;

XX oxidative metabolism; P450RAI; retinoic acid; RA; promoter; ss.

XX Danio rerio.

XX Key Location/Qualifiers

XX CDS 87..1566

XX FT /tag= a

XX FT /product= "Cytochrome zp450RAI"

XX WO9749832-A2.

XX 31-DEC-1997.

XX 23-JUN-1997; 97WO-CA00488.

XX 01-OCT-1996; 96US-0724466.

XX 21-JUN-1996; 96US-0667546.

XX (TOOH) UNIV QUEENS KINGSTON.

XX Petkovich PW;

XX WPI; 1998-077193/07.

XX P-PSDB; AAW37733.

XX Identifying DNA encoding inducible or suppressible cytochrome P450 -

XX pt screening for drugs which reduce the catabolism of retinoic acid,

XX pt useful in cancer chemotherapy and the treatment of acne and

XX psoriasis

XX Disclosure; Pages 59B-59D; 113pp; English.

XX This nucleotide sequence encodes the cytochrome zp450RAI of the

XX zebra fish. Its expression is dependent on the presence of retinoic

XX acid (RA). The retinoid-regulated genes such as the inducible

XX cytochrome P450RAI gene specifically metabolises a derivative of the

XX RA. The cytochrome P450 gene in general produces enzymes involved

XX in the oxidative metabolism of endogenous and exogenous compounds.

XX The cytochrome P450 nucleotide sequence can be used to induce or

XX suppress the expression of its protein. P450RAI is highly induced

XX by RA in cell lines and tissues. This allows for development of a

XX drug screen using promoters and nucleotide sequences to identify

XX drugs which are useful for reducing the catabolism of RA.

XX Sequence 1850 BP; 519 A; 391 C; 454 G; 486 T; 0 other;

alignment_scores:

Quality: 1738.00

Ratio: 4.080

Percent Similarity: 85.714

Percent Identity: 68.008

alignment_block:

US-09-668-482-4 x AAV09251

Align seg 1/1 to: AAV09251 from: 1 to: 1850

1 MetGlyLeuProAlaLeuAlaSerAlaLeuCysThrPheValIleuPr 17
|||||
88 ATGGGCTTACACCTTATGGTCACCTTCTGTGCACCATGCTGTACC 137
17 oleuLeuLeuPheLeuAlaAlaIleLysLeuThrPaspLeuTyrcysValS 34
|||||
138 CGTTTACTCTTTCTCCGCCGGGTGAAGTTGTGGAGATTAATGATGCC 187
34 eRGlyArgAspArgSerCysAlaLeuProLeuProGlyThrMetGly 50
|||||
188 GACGAGTCGATCCGAACTGCAGAACTCTACCGCCAGATCCATGGGC 237
51 PheProPhePheGlyGlyIleThrLeuGlnMetValIleGlnArgArgIysPh 67
|||||
238 TTGCCGTTATTCGAGAAACGCTCCAGCTGATCTCCAGAGAAAGAGATT 287
67 eLeuGlnMetLysArgArgLysTyrglyPheIleTyrglyThrHisLeuP 84
|||||
288 TCTGGCATGAAGAACGGCAGAAATACGGGTGCATCTACAGACGCACTCT 337
84 heGlyArgProThrValArgValMetGlyAlaAspAsnValArgArgIle 100
|||||
338 TCGGGAACCCGACTGTGAGGTGATGGAGCTGATTAATGTAGAGATT 387
388 CTGCTGGGCGGACACACAGCTGGTGTCTGTCACTGGCCAGCATGATGAG 437
101 LeuLeuGlyAspAspArgLeuValSerValHisTrpProAlaSerValAr 117
|||||
117 gThrIleLeuGlySerGlyCysLeuSerAsnLeuHisAspSerHisL 134
|||||
438 AACCATCTGGGGCTCTGCACCCCTCCATATGTCATGAGTCAACACA 487
134 ySGlnArgLysLysValIleMetArgAlaPheSerArgGlnAlaLeuGlu 150
|||||
488 AAACACAGAAAGAGCCATTATGAGGCGCTTCTCGAGATCTCTGGAG 537
151 CysTyrgValProValIleThrGluValGlySerSerLeuGluGlnTr 167
|||||
538 CACTACATTCCTCCGTATCCAGCAGAGGTAAGACGCCATACAGAAATG 587
167 Pleu.....SerCysGlyGlyArgGlyLeuLeuValTyrrProGluV 181
|||||
588 GCTGCAAAAAGACTCTCTC.....GTGCTGTATTATCCAGAAA 625
181 allyArgLeuMetPheArgIleAlaMetArgIleLeuLeuGlyCysGlu 197
|||||
626 TGAAGAACTCATGTTTCGATAGCTATGAGATCTCTGTTTGA 675
198 Pro...GlnLeuAlaGlyAspGlyAspSerGluGlnLeuValGluAl 213
|||||
676 CCAGAGCAAAATTAAGACGCGAC.....GAGCAAGAACTGTGGAAGC 716
213 aPheGlnIleMetThrArgAsnLeuPheSerLeuProIleAspValProp 230
|||||
717 TTTTAGGAAATGATCAAAACTGTCTCTGCAATCGAGCTTCTCT 766
230 heSerGlyLeuTyrgArgLysMetLysAlaArgAsnLeuIleHisAlaArg 246
|||||
767 TCAGTGTCTGTGACAGGGGTTTGAGCGCACCGCAATTCATTCACCCAAA 816
247 IleGluGlnAsnIleArgAlaLysIleCysGlyLeuArgAlaSerGluAl 263
|||||
817 ATTGAGGAAACATCAGAGAGAAATTCAGAT...GAGCAAAATGAAAA 863
263 aGlyGlnGlyCysLysAspAlaLeuGlnLeuLeuIleGlnHisSerTrpG 280
|||||
864 CGAACGAAATATCAAAAGACGCCCTTCAGCTGTGATCGAAGACAGCAGAA 913
280 lueArgGlyLysArgLeuAspMetGlnAlaLeuLysGlnSerSerThrGlu 296
|||||
914 GAAGTGACGAACCTTTTGTGTTGCGAGCGATGAAGAAAGACGATACAGAG 963

```

297 LeuLeuPheGlyGlyHisGluThrThrAlaSerAlaAlaThrSerLeuL 313
|||||
964 CTCTATTTGAGGATGAAACACCGCCAGCAGCTCAACCTCATTGT 1013
313 eThrTyrLeuGlyLeuTyrProHisValLeuGlnLysValArgLysL 330
|||||
1014 CATGTTTCTGGGTCTGAACACAGAGTGTGTGAGAGTGTGAGAGG 1063
330 eLysSerLysGlyLeuLeuGlyLysSerAsnGlnAspAsnLysLeuAsp 346
|||||
1064 TTCAGAGAGAAGTTGAATGGGCACTGTATACCTCGAAGAGCGCTTGAGT 1113
347 MetGluLeuLeuGlnGlnLeuLysTyrIleGlyCysValIleLysGluTh 363
|||||
1114 ATGAGAGCTGTTGGACACGCTGAGTACAGTGTGATTAAAGAGAC 1163
363 rLeuArgLeuAsnProProValProGlyGlyPheArgValAlaLeuLysT 380
|||||
1164 TCTTGAATCAACCTCTCTGTCCGGAGATTCTAGAGTCCGACATCAAAA 1213
380 hrPheGluLeuAsnGlyTyrGlnIleProLysGlyTyrPAsnValIleTyr 396
|||||
1214 CTTTGAATGATGATGTACCAAAATTCCTAAAGATGAGACGCTATTAC 1263
397 SerIleCysAspThrHisAspValAlaGluIlePheThrAsnLysGluG 413
|||||
1264 AACATCTGTGACACGACGATGTGGCCGACGCTTCCAAACAAAGAGGA 1313
413 urPheAsnProAspArgPheSerAlaProHisProGlnAspAlaSerArgP 430
|||||
1314 GTTCCAGCGCGAGAGATTATAGCAAAAGTGTGAGAGAGCGGCTCAGGT 1363
430 hSerPheIleProPheGlyGlyLeuArgSerCysValGlyLysGlu 446
|||||
1364 TTAACACTACATCCCTCGAGAGAGATCCAGATGTGTGGGCAAGAG 1413
447 PheAlaLysIleLeuLeuLysIlePheThrValGlnLeuAlaArgHis 463
|||||
1414 TTCGCCAAGGTGTACTCAAGATCTTTAGTGAAGTAAACGACGATTTG 1463
463 sAspTrpGlnLeuLeuAsnGlyProProThrMetLysThrSerProThr 480
|||||
1464 CAATGTGATTCTCTCAACGACGCCCGACATGAAGAAACAGCGCCGACTA 1513
480 aLysrProValAspAsnLeuProAlaArgPheThrHisPhe 493
|||||
1514 TTACCCAGTGGACATCTCCCTACCAATTCACGTAGTTAT 1554
seq_name: /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT:AAV12203
seq_documentation_block:
ID AAV12203 standard; cDNA; 1850 BP.
XX
AC AAV12203;
XX
DT 22-JUN-1998 (first entry)
XX
DE zebrafish retinoid metabolizing protein zp450RAI cDNA.
XX
KW Retinoid metabolizing protein; P450RAI; retinoid oxidase;
KW retinoid acid; zebrafish; inhibitor; antisenese; cancer;
KW actinic keratosis; oral leukoplakia; head tumour; neck tumour;
KW non-small cell lung carcinoma; basal cell carcinoma;
KW acute promyelocytic leukaemia; skin cancer; acne; psoriasis;
KW ichthyosis; therapy; diagnosis; screening; ss.
XX
OS Danio rerio.
XX
FH Key Location/Qualifiers
FT CDS 88..156
FT /tag= a
XX
PN M09749815-A1.

```

```

XX
PD 31-DEC-1997.
XX
PF 23-JUN-1997; 97WO-CA00440.
XX
PR 01-OCT-1996; 96US-0724466.
XX
PR 21-JUN-1996; 96US-0667546.
XX
PA (TOOH ) UNIV QUEENS KINGSTON.
XX
PI Beckett BR, Jones G, Petkovich PM, White JA;
XX
DR WPI: 1998-077178/07.
XX
DR P-PsDB; AAV12203.
XX
PT Retinoid metabolizing protein - useful to develop products to treat,
PT e.g. cancer, actinic keratosis, oral leukoplakia, acne, psoriasis or
XX ichthyosis
XX
PS Claim 15; Page 52-54; 110pp: English.
XX
XX This cDNA clone codes for a novel zebrafish retinoid metabolizing
XX protein (see AAV12203) designated zp450RAI. The encoded protein is a
XX retinoid oxidase that has the ability to hydroxylate retinoid acid
XX at the 4 position of the beta-ionone ring and is inducible in
XX epithelial cells exposed to retinoid acid. To isolate the clone,
XX transcripts present in fin tissue regenerating in the presence or
XX absence of retinoid acid were compared using a differential display
XX PCR technique, and an isolated clone (see AAV12203) was used to screen
XX a D. rerio 6-18 hr embryo cDNA library. Human (see AAV12204) and
XX mouse (see AAV12205) P450RAI clones have also been identified. The
XX isolated clones can be used in the recombinant production of
XX P450RAI proteins. Antisense nucleic acids can be used in a claimed
XX method for inhibiting retinoid acid hydroxylation for the treatment
XX of cancer, actinic keratosis, oral leukoplakia, secondary tumours of
XX the head and/or neck, non-small cell lung carcinomas, basal cell
XX carcinomas, acute promyelocytic leukaemia, skin cancer, and
XX CC premalignancy associated with actinic keratosis, acne, psoriasis
XX and/or ichthyosis. Promoter sequences (see AAV12206-08) are also
XX claimed.
XX
SQ Sequence 1850 BP; 519 A; 391 C; 454 G; 486 T; 0 other;
XX
alignment_scores:
Quality: 1738.00 Length: 497
Ratio: 4.080 Gaps: 5
Percent Similarity: 85.714 Percent Identity: 68.008
alignment_block:
US-09-668-482-4 x AAV12203
Align seg 1/1 to: AAV12203 from: 1 to: 1850
1 MetGlyLeuProAlaLeuLeuAlaSerAlaLeuGlyThrPheValLeuPr 17
|||||
88 ATGGGCGCTGACACCTTATGTGCTACCTTCTGACACATGCTCTAC 137
17 oLeuLeuLeuPheLeuAlaAlaIleLysLeuTrpAspLeuTyrCysValS 34
|||||
138 CGTTTACCTCTTCTGCGCGGTGAAGTGTGGAGAGATTAAATGATCC 187
34 eGlyLysrAspArgSerCysAlaLeuProLeuProProGlyThrMetGly 50
|||||
188 GACGAGTCAATCGAAGTCTGAGAGTCTTCCGCGAGAGTATGGGC 237
67 eLeuGlnMetLysArgArgLysTyrGlyPheIleTyrLysThrHisLeuP 84
|||||
238 TTGCGCTCATTTGAGAGAAACGCTCCAGCTGATCTCCAGAGAGAGT 287
67 eLeuGlnMetLysArgArgLysTyrGlyPheIleTyrLysThrHisLeuP 84
|||||
288 TCTGCGCATGAACGGCAGAAATACGGGTGCATCTTCAAGAGACGACCTCT 337

```

```

84  heglyArProthrValArgValMetGlyAlaAspAsnValArgArgIle 100
    |||||.....
338  TCGGAAACCCGACTCTCAGGGTGTGAGGAGCTGATATATGTGAGCGATT 387
101  LeuLeuGlyAspAspArgLeuValSerValHisTrpProAlaSerValAr 117
    |||||.....
388  CTGCTGGGCGAACAACAGCTGTCTCTTCAGTGGCCAGCATCAGTGGAG 437
117  gThrIleLeuGlySerGlyCysLeuSerAsnLeuHisAspSerHisL 134
    |||||.....
438  AACCATCTGGGCTGTGACACCTCTCCATGTCCATGTGAGTTCAACACA 487
134  ysgInArgLysLysValIleMetArgAlaPheSerArgLysAlaLeuGlu 150
    |||||.....
488  AAACACAAAGAAAGGCCATTTATGAGGGGTTCTCTCGAATCTCTGGAG 537
151  CysIlyrValProValIleThrGluGluValGlySerSerLeuGluGlnTr 167
    |||||.....
538  CACTACATTCCTCCGTGATCCAGCAGAGGTGAAGAGCCCATACAGAAATG 587
167  pLeu.....SerCysGlyGluArgGlyLeuLeuValTyrProGlu 181
    |||||.....
588  GCTGCAAAACACCTCTCC.....GTCCTGGTTTATCCAGAAA 625
181  alLysAlaGluMetCpheaArgIleAlaMetArgIleLeuLeuGlyCysGlu 197
    |||||.....
626  TGAAGAAACATCATGTTTCGATAGCTATGAGAAATCCTCTGGTTTGA 675
198  Pro...GlnLeuAlaGlyAspGlyAspSerGluGlnGlnLeuValGluAl 213
    |||||.....
676  CCAGAGCAAAATAAGACGGAC.....GACGACAGAACTGGTGAAGC 716
213  aPheGluGluMetThrArgAsnLeuPheSerLeuProIleAspValProp 230
    |||||.....
717  TTTTGAGGAAATGATCAAAACTTGTCTCTTCCTGCAATTCACGTTCTCT 766
230  heSerGlyLeuTyrArgGlyMetLysAlaArgAsnLeuIleHisAlaArg 246
    |||||.....
767  TCACTGGTCTTACAGGGGTTTGAGGGCAGCAATTCATTCACCTCCAAA 816
247  IleGluGlnAsnIleArgAlaLysIleCysGlyLeuArgAlaSerGluAl 263
    |||||.....
817  ATTGAGGAAACATCAGAGAAAGAAATTCAGAT...GACGACATGAAA 863
263  aGlyGlnGlyCysLysAspAlaLeuGlnLeuIleGluHisSerTrpG 280
    |||||.....
864  CGAACAGAAATACAAAGACGCCCTTCAGCTGTGATCAGAACACAGAAA 913
280  LuArgGlyGluArgLeuAspMetGlnAlaLeuLysGlnSerSerThrGlu 296
    |||||.....
914  GAAGTGACGAACCTTTAGTTTGCAGCGGATGAGAAAGAGCGTACAGAG 963
297  LeuLeuPheGlyGlyHisGluThrThrAlaSerAlaAlaThrSerLeuI 313
    |||||.....
964  CTTCTATTATTTGAGGTGATGAACACCGCCAGCAGCTCAACCTCACTGT 1013
313  eThrTyrLeuGlyLeuTyrProHisValIleGlnLysValArgGluGlu 330
    |||||.....
1014  CAGTTTCTGTGGTGAACACAGAAAGTGTGACAGAAAGTACAGAGAGAG 1063
330  euLysSerLysGlyLeuLeuCysLysSerAsnGlnAspAsnLysLeuAsp 346
    |||||.....
1064  TTTCAGGAGAAAGTTGAATGCGCATGTATACACCTGAGAAAGGCTTAGT 1113
347  MetGluIleLeuGluGlnLeuLysTyrIleGlyCysValIleLysGluTh 363
    |||||.....
1114  ATGGAGCTGTTGGACACAGCTGAAGTACACTGATGTGTAAAGAGAC 1163
363  rLeuArgLeuAsnProProValProGlyGlyPheArgValAlaLeuLysT 380
    |||||.....
1164  TCTTAGAATACACCTCTGCTGTCCCGAGGATTTCAGAGTGCACCTCAAAA 1213

```

```

380  hrPheGluLeuAsnGlyTyrGlnIleProLysGlyTyrAsnValIleTyr 396
    |||||.....
1214  CCTTGAATTTGAATGGTTACCAAAATTCCTAAAGATGAGAAACGTCAATTAC 1263
397  SerIleCysAspThrHisAspValAlaGluIlePheThrAsnLysGluI 413
    |||||.....
1264  AGCATCTGTGACACGACGATGTGGCCGACGCTTTCACAAACAAAGAGAA 1313
413  uPheAsnProAspArgPheSerAlaProHisProGluAspAlaSerArgp 430
    |||||.....
1314  GTTCCAGCCGAGAGATTTCATGACCAAAAGTCTGAGAGCGGCTCAGGT 1363
430  heSerPheIleProPheGlyGlyGlyLeuArgSerCysValIleLysGlu 446
    |||||.....
1364  TTAATCACTATCCCTTCGAGAGGATCCAGGATGTGTGTGGCAAAAGAG 1413
447  PheAlaLysIleLeuLeuLysIlePheThrValGluLeuAlaArgHisCy 463
    |||||.....
1414  TTGGCCAAAGTGTACTCAGATCTTTTATGTTGATTAAACGACGACATTG 1463
463  sAspTrpGlnLeuLeuAsnGlyProProThrMetLysThrSerProThrV 480
    |||||.....
1464  CAATTGATTTCTCTCAACCGACCCCGACATGAAACAGGCGCCGACTA 1513
480  alTyrProValAspAsnLeuProAlaArgPheThrHisPhe 493
    |||||.....
1514  TTTACCCAGTGGACAAATCTCCTACCAATTCACCTAGTTAT 1554
seq_name: /SIDS1/gcgdata/geneseq/geneseqn/NAI1998.DAT:AAV12216
seq_documentation_block:
ID   AAV12216 standard; DNA; 4164 BP.
XX
XX   AAV12216;
XX
XX   22-JUN-1998 (first entry)
XX
XX   Mouse retinoid metabolising protein mp450RAI genomic DNA.
XX
XX   Retinoid metabolising protein; P450RAI; retinoid oxidase;
XX   retinoic acid; mouse; inhibitor; antisense; cancer;
XX   actinic keratosis; oral leukoplakia; head tumour; neck tumour;
XX   non-small cell lung carcinoma; basal cell carcinoma;
XX   acute promyelocytic leukaemia; skin cancer; acne; psoriasis;
XX   ichthyosis; therapy; diagnosis; screening; cytochrome P450; ds.
XX
XX   Mus musculus.
XX
XX   MO9749815-A1.
XX
XX   31-DEC-1997.
XX
XX   23-JUN-1997; 97WC-CA00440.
XX
XX   01-OCT-1996; 96US-0724466.
XX   21-JUN-1996; 96US-0667546.
XX
XX   (TOOH ) UNIV QUEENS KINGSTON.
XX
XX   Beckett BR, Jones G, Petkovich PM, White JA;
XX
XX   WPI; 1998-077178/07.
XX
XX   Retinoid metabolising protein - useful to develop products to treat,
XX   e.g. cancer, actinic keratosis, oral leukoplakia, acne, psoriasis or
XX   ichthyosis
XX
XX   Disclosure: Page 69-70; 110pp; English.
XX
XX   This sequence comprises a genomic clone of a novel mouse retinoid
XX   metabolising protein designated mp450RAI. A cDNA clone (see
XX   AAV12205) that includes a coding sequence for the full-length
XX   mp450RAI protein (see AAV4161) is also provided. The encoded
XX

```

CC protein is a retinoid oxidase that has the ability to hydroxylate
CC retinoic acid at the 4 position of the beta-ionone ring and is
CC retinoic acid in epithelial cells exposed to retinoic acid. Zebrafish,
CC inducible in epithelial cells exposed to retinoic acid. are claimed.
CC human and mouse P450RAI cDNA clones (see AAV12203-05) are claimed.
CC They can be used in the recombinant production of P450RAI.
CC Antisense sequences can be used in a claimed method for inhibiting
CC retinoic acid leukoxylatation for the treatment of cancer, actinic
CC keratosis, oral leukoplakia, secondary tumours of the head and/or
CC neck, non-small cell lung carcinomas, basal cell carcinomas, acute
CC promyelocytic leukaemia, skin cancer, and premalignancy associated
CC with actinic keratosis, acne, psoriasis and/or ichthyosis. Promoter
CC sequences (see AAV12206-08) for the P450RAI genes are also claimed.

Sequence 4164 BP; 997 A; 975 C; 1071 G; 1121 T; 0 other;

alignment_scores:

Quality: 17/12.00	Length: 1048
Ratio: 4.019	Gaps: 10

Percent Similarity:	40.727	Percent Identity:	38.815
---------------------	--------	-------------------	--------

```
alignment_block:
```

Align seg 1/1 to: AAV12216 from: 1 to: 4164

40	CyAlAlaLeuProLeuProProGlyTThMeCylPheProPhePhGlyYL	56
:		99
73	TGGGGTCTCTCTCCACCTCCCTC	99
56	uThLeuGlnMeValLeuGlnArGArGylsPheLeuGlnMeLysArgA	73
		73
100CAGCGAGGAAGATTCTGCAGATGAAGCCCA	130
73	rgLysTyrGlyPheIleTyrTyrThrHisLeuPheGlyArGProthrvAl	89
		89
131	GGAATTCGGCTTCATCTACAAAGCGCATCTGTGTGGGGGCCACGGGAG	180
90	ArgValMeGlyAlaAspAsnValArGArgIleLeuLeuGlyAspAspAr	106
		106
181	CGGGTATGGCGCGGGATATGTGGCGGCATCTGTGGAGAGCACCg	230
106	gLeuValSerValHisTrpProAlaSerValArGhrIleLeuGlySerG	123
		123
231	GTTCGTTCGGTGGCACTGGCCCCGCTCGGTGGGCACCACTCGGGGCGT	280
123	LycSLeuSerAsnLeuHisAspSerSerHisLysGlnArGyls..Lys	138
		138
281	GCTGCTCTCCAACTGCACGATTCTCGCACACGCGGAAGAAGGTG	330
138	138
331	AGGGTGAGCTGGCAACTCCTTGGCTGGCAGGAGACCTCATCCTATGGCT	380
138	138
381	TGGTTCAGGCAAAATAGATCGGGGCCAGGGCTAGTCTATGTGGTGG	430
138	138
431	GACCAGGACCTCTATATCTGATGCCACTTAGCTTTCTGCTAGCAGC	480
138	138
481	TGGGTTAGTCTCGGGGGGAGCTGAATTCTTGAAAGGTACTGGAAAG	530
138	138
531	CGAAGGGGGGGGGCGTAGGGAAGTAGAGATTGTAACACTGCTGCTC	580
139ValIleMetArGAlaPheSerArgLAlaLeuG	150
		150
581	CTGGGGGGTCTCAGGTGATTATGACAGGCTTACGGCGCAGGCACTTCCA	630

150	ucstyTyaValProValIlethrgluInuValGlyserIleuInuInt	167
151	gTgcTACCTGCCCTGATCCCTAGAGAACTACAGACTTGTCTGGACACT	680
152	rPLeuSerCysGlyGluArgGlyLeuLeuValTyrProGluValIlySarG	183
153	661 GGCCTAGCTGGGGGACCGCGGCTCTGGTCTTACCCCGAGGTGAAGCCG	730
154	LeuMetPheArgIleAlaMetArgIleLeuLeuGlyCysGluProGluIle	200
155	731 CTCATGTTCCCATCCGCGCATGCCGATCTCTGGCTGGCAGCCGGGCTCC	780
156	200 uAlaGlyAspGlyAspSerGluGlnGlnLeuValAlaGluAlaPheGluGluM	217
157	781 ACCGGGCGCGGGGAGGAGAGACAGACAGCACTGTGTGGAGGCTTTGGAGAGA	830
158	217 eLThrArgAsnLeuPheSerLeuProIleAspValProPheSerGlyLeu	233
159	831 TGACCCGCAATCTCTCTCTCTCCATTGACGTGGCCCTTTACCGGCTTG	880
160	234 TyrArg.....	235
161	881 TACCGGGTAAAGGCGGCTTTCGGAGTGGAGTACGGGAACCGAAGCTCGG	930
162	236Gly.MetIlyAlaArg	240
163	931 GCATCCGCTCACCGGCACAGCTCTCTCCGCGCTCAGGGCGTGAAGAGCGCGG	980
164	241 AsnLeuIleHisAlaArgIleGluGlnAsnIleArgAlaIlyIleCysGlu	257
165	981 AACCTTATACAGCGCGGATCGAGGAGAACCTTGCAGCCCAAGATCGCGCG	1030
166	257 yLeuArgAlaSerGluAlaGlyGlnGlyCysIlyAspAlaLeuGlnLeuL	274
167	1031 GCTTCAGGCTACAGAGCCGAGTGGGGGTGTGAAGAGCGCGTGCAGCTCC	1080
168	274 euIleGluHisSerTyrPheLysArgGlyGluArgLeuAspMetGln.....	288
169	1081 TGAATTAGCACTCGTGGGAGAGGGGAGAGAGCTGTGATATGCAGAGTGA	1130
170	288	288
171	1131 AGCAATTTCAAAAGGTGCCAAGGGCCGGGAGGTGCTCTGACTTTCCAGA	1180
172	288	288
173	1181 CACACTTCTGGGGGTCTCCAAAGCCCTGTCAAGGCCCGACTACTTCCAA	1230
174	288	288
175	1231 GTGGGCGGCGATGCTAGTGTAGAGCTTTTCAACCTGTGGGTCTGAACC	1280
176	288	288
177	1281 CTTACAGGAGCCAAACAACCTTTTCAAGAAAGGTGCGCTAAGAGCATCTGC	1330
178	288	288
179	1331 ATATCCGATTATTACATCAAGAAACATACAGTAGCAAAATTTACCGTTAT	1380
180	288	288
181	1381 GAGGTAGCAACAAGATTAATTTTATCTGTTGGGGGTCCACCACAACAGGAG	1430
182	288	288
183	1431 AACGCTATTAAAGGTGGGCACTTGTCTAGAGAGCTGTGAAGGGGGTGGC	1480
184	288	288
185	1481 TGAACAATTGGGGAAGATCCCAAAAGTTCAAAAGGCAAGGCTCATTTACAA	1530

```

288 ..... 288
1531 GGTAAAGGGAAGACGATTAAAGGAGTTTGCGTTTGTGTGG 1580
288 ..... 288
1581 TCTTGACTTCTATGAAACAAACGGAATTTTACCTTGAAGTCTCCGTG 1630
288 ..... 288
1631 CAATATTCAGGTGAGTCTTTGTAACAGTGTATTAACGACTCAGA 1680
289 ..... AlaLeuylsGlnSerSerThr 295
1681 TCTGTATAAACTTCGTTTTTATTCCTTAGGCACTAAACATGCTCAACA 1730
296 GluLeuLeuPheGlyGlyHisGlnThrAlaSerAlaAlaThrSerLe 312
1731 GAGCTCCTCTTGGTGGTCTATGAAACTACAGCCAGTGTGCGACGCACT 1780
312 uLeuThrTyrLeuGlyLeuTyrProHisValLeuGlnLysValArgGlu 329
1781 GATCACTTAACCTAGAGACTACCCACATGTCTCCAGAAAGTTGAGAG 1830
329 LuLeuLysSer ..... 332
1831 AGATAAAGAGCAAGTAGGATGATTCTAGAGGTGCCATTTGCCCTAGA 1880
332 ..... 332
1881 CATTCCTATTAACCCACCACCACCCACCTGTATATAAGTTTGTCTC 1930
332 ..... 332
1931 GATACACCCAGTACTATGACAGTGAAGATCTGAGAGTAGTGGGACTGT 1980
332 ..... 332
1981 GGGGAGAGACTCCACCTCGTAATTTAAAAAGCAGTTGTTGTACTGG 2030
332 ..... 332
2031 GCTCTCTCTGGGAGAAATTGACCCCTCTCCTCTCTCTCTCTCTCTCT 2080
332 ..... 332
2081 CCTCTCTCTCTCCACACACACACCATCACCCCTTTTATAGAGCAAG 2130
332 ..... 332
2131 TTCTCCTTCCCTGACCAAGAACAATGAATATGTGATTAGAGCCAAATAG 2180
332 ..... 332
2181 TGATCAGGAGTGCAGTGTGGTGAAGGCTCAGGGATAGACCTTTATATA 2230
332 ..... 332
2231 CCTGATTAAGCAACATTTGCTGATTAATGGTTTAGGCTGAGAGAGTGTG 2280
332 ..... 332
2281 AAAGAGAGGCCATCAGGCCATCAGCTCTTTCCCTTTTATCTCTCCATC 2330
332 ..... 332
2331 CAGAGCGCTTCAGGTTAGTTAACAGAGTGAAGTCTGCTGGGCTGACTTT 2380
333 ..... Lys. 333
2381 TTTTGGAGTCCAGGATCCACTACTCTTTTATCTGTTCCATA 2430
334 GlyLeuLeuLysSerAsnGlnAspAsnLysLeuAspMetGluLeu 350

```

```

2431 GGGCTTACTTTGCAAGAGCAATCAGACAAACAAGTTTACATGGAACCT 2480
350 euGlnLeuLysTyrIleGlyCysValIleLysGlnThrLeuArgLeu 366
2481 TGACACAGCTTAAATACACTGGGTGTGTCATTTAAGAGAACCCGCGATTG 2530
367 AsnProProValProGlyGlyPheArgValAlaLeuLysThrPheGlu 382
2531 AATCTCGGTTCCAGAGAGGTTTGGGTTGCTGTGAAGACTTTTGAGCT 2580
382 ..... 382
2581 GAATGTAGTGCACCTCTGTCCCCACCCACCCAGCCCTGTGCAGCTCCA 2630
382 ..... 382
2631 CTCTGCTATGCTGTGAGCATCAGTGCACAGACAGTGGCTCATGCCC 2680
382 ..... 382
2681 TTGACAGTCTCTGCTCCTATAGTACTGGAAACCAATTTGCTTCCTCT 2730
382 ..... 382
2731 CTTAATGCCATCATGCTAGTAATGACTTTTGTGTGCAAGCTCAGGG 2780
383 ..... LeuAsn..GlyTyrG 387
2781 CCGGATTTGTCAAATCTTAGATTTTTTTTTTTTTTAAACAGGGAATCC 2830
387 InIleProLysGlyTrpAsnValIleTyrSerIleCysAspThrHisAsp 403
2831 AGATCCCAAGGCGTGGAAATGTTATTACAGTATCTGTACACCCACGAT 2880
404 ValAlaGluIlePheThrAsnLysGlnIlePheAsnProAspArgPhe 420
2881 GTGGCAGATATCTTCATCAACAAGAGGAAATTAATCCGACCGCTTAT 2930
420 rAlaProHisProGluAspAlaSerArgPheSerPheIleProPheGly 437
2931 AGTGCCTATCCAGAGAGATGCTTCCGGTTCAGCTTCATTCATTGGAG 2980
437 IlyIleuArgSerCysValGlyGlnPheAlaLysIleLeuLeuLys 453
2981 GAGGCCCTTCGAGCTGTGTAGGCAAAAGATTGCAAAAATCTCTTAG 3030
454 IlePheThrValGluLeuAlaArgHisCysAspTrpGlnLeuLeuAsn 470
3031 AATTTACAGTGAAGCTGCTAGGCACTGTGATGGCAGCTTCTTAAATGG 3080
470 yProProThrMetLysThrSerProThrValTyrProValAspAsnLeu 487
3081 AACTCTCAATGAAGACAGCCACACTGTGTACCTGTGACCAATCTCC 3130
487 roAlaArgPheThrHisPheHisGlyIle 497
3131 CTGCAAGATTTAACCCACTTCCAGGAGATATC 3162
seq_name: /SIDSL/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV12214
seq_documentation_block:
ID AAV12214 standard; DNA: 2677 BP.
XX
AC AAV12214;
XX
DT 22-JUN-1998 (first entry)
XX
DE Human retinoid metabolising protein hp450RAI genomic DNA.
XX
KW Retinoid metabolising protein; P450RAI; retinoid oxidase;
KW retinoic acid; human; inhibitor; antisense; cancer;
KW actinic keratosis; oral leukoplakia; head tumour; neck tumour;

```


KW non-small cell lung carcinoma; basal cell carcinoma;
 KW acute promyelocytic leukaemia; skin cancer; acne; psoriasis;
 KW ichthyosis; therapy; diagnosis; screening; cytochrome P450; ds.
 XX Homo sapiens.
 XX MO9749815-A1.
 XX
 PD 31-DEC-1997.
 XX
 PF 23-JUN-1997; 97MO-CA00440.
 XX
 PR 01-OCT-1996; 96US-0724466.
 PR 21-JUN-1996; 96US-0667546.
 XX
 PA (TOOH) UNIV QUEENS KINGSTON.
 PI Beckett BR, Jones G, Petkovich PM, White JA.
 DR WPI; 1998-077178/07.
 XX
 PT Retinoid metabolising protein - useful to develop products to treat,
 PT e.g. cancer, actinic keratosis, oral leukoplakia, acne, psoriasis or
 PT ichthyosis
 XX
 PS Disclosure: Page 68; 110pp; English.
 XX
 CC This sequence comprises a genomic clone of a novel human retinoid
 CC metabolising protein designated hp450RAI. Another genomic clone
 CC (see AAV12215) is also provided, as well as a cDNA clone (see AAV12204)
 CC that includes a coding sequence for the full-length hp450RAI
 CC protein (see AAW44160). The encoded protein is a retinoid oxidase
 CC that has the ability to hydroxylate retinoic acid at the 4 position
 CC of the beta-ionone ring and is inducible in epithelial cells
 CC exposed to retinoic acid. The hp450RAI gene has been localised to
 CC 10q23-24. Zebrafish, human and mouse P450RAI cDNA clones (see
 CC AAV12203-05) are claimed. They can be used in the recombinant
 CC production of P450RAI. Antisense nucleic acids can be used in a
 CC claimed method for inhibiting retinoic acid hydroxylation for the
 CC treatment of cancer, actinic keratosis, oral leukoplakia, secondary
 CC tumours of the head and/or neck, non-small cell lung carcinomas,
 CC basal cell carcinomas, acute promyelocytic leukaemia, skin cancer,
 CC and premalignancy associated with actinic keratosis, acne,
 CC psoriasis and/or ichthyosis. Promoter sequences (see AAV12206-08)
 CC for the P450RAI genes are also claimed.
 XX
 SQ Sequence 2677 BP; 489 A; 775 C; 868 G; 511 T; 34 other;
 alignment_scores:
 Quality: 1365.00 Length: 669
 Ratio: 4.320 Gaps: 7
 Percent Similarity: 47.235 Percent Identity: 46.936
 alignment_block:
 US-09-668-482-4 x AAV12214 ..
 Align seg 1/1 to: AAV12214 from: 1 to: 2677
 1 MetGlyLeuProAlaLeuLeuAlaSerAlaLeuGlyThrPheValLeuPr 17
 676 ATGGGGCTCCCGGGGCTGCTGGCCAGTGCCTTGCACTTCGATGCC 725
 17 OleuleuLeuPheLeuAlaAlaIleValLeuTrpAspLeuTrpCysValS 34
 726 GCTGCTGCTTCTCCGCTGCGATCAAGCTCTGGACCTGATACGCGGA 775
 34 erGlyArgAspArgSerCysAlaLeuProLeuProGlyThrMetGly 50
 776 GCGGGCCGACCGCAGTGTGTGCTCCCATTCGCCCGGGGACTATSGS 825
 51 PheProPhePheGlyGluThrLeuGlnMetValLeu.Gln..... 63
 |||||||

826 TTCCCTTCTTTGGGAAACCTTGAGATGNTACTNCAGTAAGGAGGG 875
 63 63
 876 TGGGCGGACAGAGCTGCTTCCCGAGACCCGCGCGGCTCTGGGCTTCT 925
 63 63
 926 GCTGAATCGGGGTAGCGCCCGGGAGGCGATGCTATTGGCGCTAGAG 975
 63 63
 976 CAGGGCTGGCGGAGCGGGCGCTCCCGGKMYKSCSTCAWGCSCRCWVK 1025
 64 ArgArglyPheLeuGlnMetIlyArg 72
 1026 TMMCTCCGCTTCTTCCCAAGCGGARSAAANKCYKMRGATGAAGGCG 1075
 73 ArgLySTyrGlyPheIleTyrIleThrHisLeuPheGlyArgProThrVa 89
 1076 AGGAATACGCTTCTATCTACAGACGATCTGTGGGGCGGCCACCGCT 1125
 89 lArgValMetGlyAlaAspAsnValArgArgIleLeuLeuGlyAspAspA 106
 1126 ACGGCTGATGGCGCGGACAATGTCGGCGCATCTTGTCTCGAGAGACC 1175
 106 rGLeuValSerValHisTrpProAlaSerValArgThrIleLeuGlySer 122
 1176 GCTGTGTGCTGCTCCATGCGGCGGCTGCTGCTGCTGCTGCTGCTGCT 1225
 123 GlyCysLeuSerAsnLeuHisAspSerSerHisIlySGlnArgIlylys.. 138
 1226 GCGTGTCTTCTTAACCTGACGACATCTCTGACAAAGACGCAAGAGAGT 1275
 138 138
 1276 GGGGCGAGAGGCGAGCGCTGGACAGGAGGGGACCCCATTTATGAGCG 1325
 138 138
 1326 GAATTCGGCTGATGATGATCTAGCGCGGGCTAGCAGCTTGAAGTGGGT 1375
 138 138
 1376 AGGACCTCTGCGACGCTCCAGTTAGCTTCCAGCTCGAGAGTGCAT 1425
 138 138
 1426 GTGCTGCGAGAGCTGGGGGTGTGTGAAAGGAGCGCGGTAGACGAGAG 1475
 139 YAlle 140
 1476 GGGCGATGAGAGCTTTAAGCTGTCCCTCTCGGAGCTGAGGTGATT 1525
 141 MetArgAlaPheSerArgGluAlaLeuGluCysTyrValProValIleIle 157
 1526 ATCGGGCTTACAGCCGCGAGGCACTCGAATGCTACGTCCCGGTGATCAC 1575
 157 rGluGluValAlaGlySerSerLeuGluGlnTrpLeuSerCysGlyGluArg 174
 1576 CGAGGAAGTGGGAGAGCTCTGAGAGAGTGGCTGAGCTGGCGGAGCGGG 1625
 174 lLeuLeuValTyrProGluValIlyArgLeuMetPheArgIleAlaMet 190
 1626 GCTGCTGCTTACCCCGGAGGTGAAGGCTCATGTCTCCAAATCGCCATG 1675
 191 ArgIleLeuLeuGlyCysGluProGlnIleAlaGlyAspGlyAspSerG 207
 1676 CGCATCTTACTGGCTGCGAACCCTGCGGCGGCGGAGGACTCTCGA 1725
 207 uGlnGlnLeuValGluAlaPheGluGluMetThrArgAsnLeuPheSerL 224
 1726 GCACGACTTGTGGAGGCTTCAGAGAAATGACCGGCAATCTCTTCCG 1775

```
224 euProIIaspValProPheSerGlyLeuTyArg..... 235
|||||
1776 TGCCATGAGAGTGCCCTTCAGCGGGCTGTACCGGTAAGGGCGGCAAC 1825
235 ..... 235
1826 GGGCTGCGAGCTAGGGGCGCGGACCTGGCGGCTCTACCGCGCGCG 1875
236 .....Gly.MetLysAlaArgAsnLeuIleHisAlaArgIle 247
|||||
1876 CTCCTGCGGCTCAGGCGCATGAAAGCGCGCAACCTCATTCACGCGCATC 1925
248 GluGlnAsnIleArgAlaLysIleCysGlyLeuArgAlaSerGluAlaG1 264
|||||
1926 GAGCAGACATTCGCGCCAGATCTGCGGGCTGCGGGCATCCGAGCGGG 1975
264 yGIuGIyCysLysAspAlaLeuGlnLeuLeuIleGluHisSerTrpGluA 281
|||||
1976 CCAGGGCTGCAAAAGCGCGCTGCACACTGTGTATCGAGCAGCATCTGGAGA 2025
281 rGgLYgluArgIleuAspMetGln..... 288
|||||
2026 GGGAGAGAGCGGCTGGACATGCAAGTGAGTAGCAGCTTACAGCAGGCACT 2075
288 ..... 288
2076 GCGAGATTGTCCTCCCTGCTTCCAAAGCGCTGTCTCTGGGGCCCCCAA 2125
288 ..... 288
2126 AGCGCGCGCTGGGGCCAGCTTCTTGAGTGGGCGCGCGCTCAGACTA 2175
288 ..... 288
2176 CAGCTATGATCCCGAAGAAGGTGAGACACCCGTCAGAGAGCTGC 2225
288 ..... 288
2226 GGAAGGGCTCGGAGAACCTGGAGCATCCCTAGCCTTAMCAGGTT 2275
288 ..... 288
2276 TCAAGAGGAAGTTGGAATTGCAGAAAATGTTAATAAGAACCTTGCAT 2325
288 ..... 288
2326 TTTAATAAACTAGACTTTAATCAGAGATTCCGTAAGCGGGGCTGC 2375
288 ..... 288
2376 TACTCGCTTACTGCTCCAGCTGAACATAAGGAGCGTTCATTTGTTTA 2425
288 ..... 288
2426 AAGATATTGCTTCCCTGACTTCTGTACAGCAAAAACATTTAGCCCTCTA 2475
288 ..... 288
2476 GTCTTCCTCCAGAACTCTCAGTTCGATTTGAGTAATCCTCTGTCAAA 2525
289 .....Ala..Leu 290
2526 CCGGAGCAGACTTGTAGAAATGTGGCTCTCAGCTCTATTTAGGCACTA 2575
291 LysGlnSerSerThrIleuLeuPheGlyGlyHisGluThrAlaIase 307
|||||
2576 AAGCAATCTTCAACGAACTCCTCTTGGAGACACGAAACACAGGCGCAG 2625
307 rAlaAlaIThrSerIleuIleThrTyLeuGlyLeuTyProHisValLeuG 324
|||||
2626 TGCAGCCACATCTGTGATCACTTACCTGGGGCTCTACCCACATGTTCGC 2675
PR
```

```
324 In 324
||
2676 AG 2677

seq_name: /SIDSI/gcgdata/geneseq/geneseqn/NA2000.DAT:AA39029
seq_documentation_block:
ID AAC39029 standard; DNA; 1685 BP.
XX
AC AAC39029;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment seq ID NO: 23101.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
```

```
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-013987.
PR 22-JUN-1999; 99US-013989.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140921.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144633.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147416.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148365.
PR 13-AUG-1999; 99US-0148684.

PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151086.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 14-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160778.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

alignment_scores:
  Quality: 478.00      Length: 490
  Ratio: 1.615        Gaps: 13
Percent Similarity: 60.408  Percent Identity: 27.551

alignment_block:
US-09-668-482-4 x AAC39029  ..
Align seg 1/1 to: AAC39029  from: 1 to: 1685
15 ValLeuPro...LeuLeuLeuAlaAlaAlaLeuTrpAspLe 30
```

```

142 GTTGTACCAATCTCATCTTATGCTTGTGTTGTAAGATG..... 183
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999

```

```

967 ACAGCTAGTTGCTTAAGTCTTGAATTTCTTAAGTACTTA.....CA 1004
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1045
1046
1047
1048
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1070
1071
1072
1073
1074
1075
1076
1077
1078
1079
1080
1081
1082
1083
1084
1085
1086
1087
1088
1089
1090
1091
1092
1093
1094
1095
1096
1097
1098
1099
1100
1101
1102
1103
1104
1105
1106
1107
1108
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1120
1121
1122
1123
1124
1125
1126
1127
1128
1129
1130
1131
1132
1133
1134
1135
1136
1137
1138
1139
1140
1141
1142
1143
1144
1145
1146
1147
1148
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1180
1181
1182
1183
1184
1185
1186
1187
1188
1189
1190
1191
1192
1193
1194
1195
1196
1197
1198
1199
1200
1201
1202
1203
1204
1205
1206
1207
1208
1209
1210
1211
1212
1213
1214
1215
1216
1217
1218
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1229
1230
1231
1232
1233
1234
1235
1236
1237
1238
1239
1240
1241
1242
1243
1244
1245
1246
1247
1248
1249
1250
1251
1252
1253
1254
1255
1256
1257
1258
1259
1260
1261
1262
1263
1264
1265
1266
1267
1268
1269
1270
1271
1272
1273
1274
1275
1276
1277
1278
1279
1280
1281
1282
1283
1284
1285
1286
1287
1288
1289
1290
1291
1292
1293
1294
1295
1296
1297
1298
1299
1300
1301
1302
1303
1304
1305
1306
1307
1308
1309
1310
1311
1312
1313
1314
1315
1316
1317
1318
1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1340
1341
1342
1343
1344
1345
1346
1347
1348
1349
1350
1351
1352
1353
1354
1355
1356
1357
1358
1359
1360
1361
1362
1363
1364
1365
1366
1367
1368
1369
1370
1371
1372
1373
1374
1375
1376
1377
1378
1379
1380
1381
1382
1383
1384
1385
1386
1387
1388
1389
1390
1391
1392
1393
1394
1395
1396
1397
1398
1399
1400
1401
1402
1403
1404
1405
1406
1407
1408
1409
1410
1411
1412
1413
1414
1415
1416
1417
1418
1419
1420
1421
1422
1423
1424
1425
1426
1427
1428
1429
1430
1431
1432
1433
1434
1435
1436
1437
1438
1439
1440
1441
1442
1443
1444
1445
1446
1447
1448
1449
1450
1451
1452
1453
1454
1455
1456
1457
1458
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1480
1481
1482
1483
1484
1485
1486
1487
1488
1489
1490
1491
1492
1493
1494
1495
1496
1497
1498
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509
1510
1511
1512
1513
1514
1515
1516
1517
1518
1519
1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1560
1561
1562
1563
1564
1565
1566
1567
1568
1569
1570
1571
1572
1573
1574
1575
1576
1577
1578
1579
1580
1581
1582
1583
1584
1585
1586
1587
1588
1589
1590
1591
1592
1593
1594
1595
1596
1597
1598
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1610
1611
1612
1613
1614
1615
1616
1617
1618
1619
1620
1621
1622
1623
1624
1625
1626
1627
1628
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1660
1661
1662
1663
1664
1665
1666
1667
1668
1669
1670
1671
1672
1673
1674
1675
1676
1677
1678
1679
1680
1681
1682
1683
1684
1685
1686
1687
1688
1689
1690
1691
1692
1693
1694
1695
1696
1697
1698
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1710
1711
1712
1713
1714
1715
1716
1717
1718
1719
1720
1721
1722
1723
1724
1725
1726
1727
1728
1729
1730
1731
1732
1733
1734
1735
1736
1737
1738
1739
1740
1741
1742
1743
1744
1745
1746
1747
1748
1749
1750
1751
1752
1753
1754
1755
1756
1757
1758
1759
1760
1761
1762
1763
1764
1765
1766
1767
1768
1769
1770
1771
1772
1773
1774
1775
1776
1777
1778
1779
1780
1781
1782
1783
1784
1785
1786
1787
1788
1789
1790
1791
1792
1793
1794
1795
1796
1797
1798
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1810
1811
1812
1813
1814
1815
1816
1817
1818
1819
1820
1821
1822
1823
1824
1825
1826
1827
1828
1829
1830
1831
1832
1833
1834
1835
1836
1837
1838
1839
1840
1841
1842
1843
1844
1845
1846
1847
1848
1849
1850
1851
1852
1853
1854
1855
1856
1857
1858
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1870
1871
1872
1873
1874
1875
1876
1877
1878
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1889
1890
1891
1892
1893
1894
1895
1896
1897
1898
1899
1900
1901
1902
1903
1904
1905
1906
1907
1908
1909
1910
1911
1912
1913
1914
1915
1916
1917
1918
1919
1920
1921
1922
1923
1924
1925
1926
1927
1928
1929
1930
1931
1932
1933
1934
1935
1936
1937
1938
1939
1940
1941
1942
1943
1944
1945
1946
1947
1948
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030
2031
2032
2033
2034
2035
2036
2037
2038
2039
2040
2041
2042
2043
2044
2045
2046
2047
2048
2049
2050
2051
2052
2053
2054
2055
2056
2057
2058
2059
2060
2061
2062
2063
2064
2065
2066
2067
2068
2069
2070
2071
2072
2073
2074
2075
2076
2077
2078
2079
2080
2081
2082
2083
2084
2085
2086
2087
2088
2089
2090
2091
2092
2093
2094
2095
2096
2097
2098
2099
2100
2101
2102
2103
2104
2105
2106
2107
2108
2109
2110
2111
2112
2113
2114
2115
2116
2117
2118
2119
2120
2121
2122
2123
2124
2125
2126
2127
2128
2129
2130
2131
2132
2133
2134
2135
2136
2137
2138
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2149
2150
2151
2152
2153
2154
2155
2156
2157
2158
2159
2160
2161
2162
2163
2164
2165
2166
2167
2168
2169
2170
2171
2172
2173
2174
2175
2176
2177
2178
2179
2180
2181
2182
2183
2184
2185
2186
2187
2188
2189
2190
2191
2192
2193
2194
2195
2196
2197
2198
2199
2200
2201
2202
2203
2204
2205
2206
2207
2208
2209
2210
2211
2212
2213
2214
2215
2216
2217
2218
2219
2220
2221
2222
2223
2224
2225
2226
2227
2228
2229
2230
2231
2232
2233
2234
2235
2236
2237
2238
2239
2240
2241
2242
2243
2244
2245
2246
2247
2248
2249
2250
2251
2252
2253
2254
2255
2256
2257
2258
2259
2260
2261
2262
2263
2264
2265
2266
2267
2268
2269
2270
2271
2272
2273
2274
2275
2276
2277
2278
2279
2280
2281
2282
2283
2284
2285
2286
2287
2288
2289
2290
2291
2292
2293
2294
2295
2296
2297
2298
2299
2300
2301
2302
2303
2304
2305
2306
2307
2308
2309
2310
2311
2312
2313
2314
2315
2316
2317
2318
2319
2320
2321
2322
2323
2324
2325
2326
2327
2328
2329
2330
2331
2332
2333
2334
2335
2336
2337
2338
2339
2340
2341
2342
2343
2344
2345
2346
2347
2348
2349
2350
2351
2352
2353
2354
2355
2356
2357
2358
2359
2360
2361
2362
2363
2364
2365
2366
2367
2368
2369
2370
2371
2372
2373
2374
2375
2376
2377
2378
2379
2380
2381
2382
2383
2384
2385
2386
2387
2388
2389
2390
2391
2392
2393
2394
2395
2396
2397
2398
2399
2400
2401
2402
2403
2404
2405
2406
2407
2408
2409
2410
2411
2412
2413
2414
2415
2416
2417
2418
2419
2420
2421
2422
2423
2424
2425
2426
2427
2428
2429
2430
2431
2432
2433
2434
2435
2436
2437
2438
2439
2440
2441
2442
2443
2444
2445
2446
2447
2448
2449
2450
2451
2452
2453
2454
2455
2456
2457
2458
2459
2460
2461
2462
2463
2464
2465
2466
2467
2468
2469
2470
2471
2472
2473
2474
2475
2476
2477
2478
2479
2480
2481
2482
2483
2484
2485
2486
2487
2488
2489
2490
2491
2492
2493
2494
2495
2496
2497
2498
2499
2500
2501
2502
2503
2504
2505
2506
2507
2508
2509
2510
2511
2512
2513
2514
2515
2516
2517
2518
2519
2520
2521
2522
2523
2524
2525
2526
2527
2528
2529
2530
2531
2532
2533
2534
2535
2536
2537
2538
2539
2540
2541
2542
2543
2544
2545
2546
2547
2548
2549
2550
2551
2552
2553
2554
2555
2556
2557
2558
2559
2560
2561
2562
2563
2564
2565
2566
2567
2568
2569
2570
2571
2572
2573
2574
2575
2576
2577
2578
2579
2580
2581
2582
2583
2584
2585
2586
2587
2588
2589
2590
2591
2592
2593
2594
2595
2596
2597
2598
2599
2600
2601
2602
2603
2604
2605
2606
2607
2608
2609
2610
2611
2612
2613
2614
2615
2616
2617
2618
2619
2620
2621
2622
2623
2624
2625
2626
2627
2628
2629
2630
2631
2632
2633
2634
2635
2636
2637
2638
2639
2640
2641
2642
2643
2644
2645
2646
2647
2648
2649
2650
2651
2652
2653
2654
2655
2656
2657
2658
2659
2660
2661
2662
2663
2664
2665
2666
2667
2668
2669
2670
2671
2672
2673
2674
2675
2676
2677
2678
2679
2680
2681
2682
2683
2684
2685
2686
2687
2688
2689
2690
2691
2692
2693
2694
2695
2696
2697
2698
2699
2700
2701
2702
2703
2704
2705
2706
2707
2708
2709
2710
2711
2712
27
```

PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130549.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132853.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-01344768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138099.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142820.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.

PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151303.
PR 30-AUG-1999; 99US-0151308.
PR 31-AUG-1999; 99US-0151308.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.

```

PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160960.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

alignment_scores:

Quality:	450.50	Length:	490
Ratio:	1.592	Gaps:	15
Percent Similarity:	57.755	Percent Identity:	27.755

alignment_block:

US-09-668-482-4 x AAC48714 ..

Align seg 1/1 to: AAC48714 from: 1 to: 1398

```

1 MetGlyLeuProAlaLeuLeuAlaSerAlaLeuCysThrPheValLeuPr 17
  |||||
22 ATGGCTCTTCTTGTATCATCGTCTCT..CTCTGT..... 54
17 OleuLeuLeuPheLeuAlaAlaIleLysLeuTrpAsp.....LeuTyrC 32
  ::|||::
55 .....TCCGCTCTCTTCGATGATCATGATGCATATATA 88
32 yValSerGlyAArgAspArgSerCysAlaLeuProLeuProProGlyThr 48
  ::|||
89 CCAGAAATGGT.....CTTCTCTCTGGAACC 114
49 MetGlyPheProPheGlyGlyLeuThrLeuGlnMetValLeuGlnArgAr 65
  |||||
115 ATGGGCTGGCCAAATCTTGGCGAAACACCGAGTTTCTCAAAACAGGCC 164
65 GlyPheLeuGlnMetLysArgArgLysTyrGlyPheIleTyrIleThr 82
  ::|||::
165 CAACCTTCATGGAAACCAAGACTCCGATAGCGAGTTCTTCAAAATCTC 214
82 ILeuPheGlyAArgProThrValArgValMetGlyAlaAspAsnValArg 98
  |||||
215 ATCTTCTAAGTTGTCACAAAGTTAATCTCAATGACTCAGAAAGTAAACGA 264
99 ArgIleLeuLeuGlyAspAspArgLeuValSerValIleThrProAlaSe 115
  |||||
265 TACATTTTAAAGACGAATCAAAAGTTGGTTGCTCGTTACCCACATC 314
115 rValArgThrIleLeuGlySerGlyCysLeuSerAsnLeuHisAspSerS 132
  ::|||
315 GATGCTTGATTAATCTTGAGACTTGTAAATGATGCGGCTTACCGGTTTCA 364

```

```

132 erHisLysGlnArgLysValIleMetArgAlaPheSerArgGluAla 148
  |||||
365 GCCACCGCTTATGAGAGCGCTCTCTCTCATATACCTGACCATG 414
149 LeuGlyCysTyrVal...ProValIleThrGluGlyValGlySerSer 164
  ::|||
415 ATGACAGATCATATCTTCCCTAAAGTTGATGACTTATCATGAAAGCTATCT 464
164 uGlnGlnTrpLeuSerCysGlyLysArgLysLeu...ValTyrProG 180
  ::|||
465 TGATCAGTGG.....AATGACCTTGAGTTATGATATCAAGATTA 505
180 lValValLysArgLeuMetPheArgIleAlaMetArgIleLeuLeuGlyCys 196
  |||||
506 AGACCAACATATGCGCATTT.....TTATCTTCA 534
197 GluProGlnLeuAlaGlyAspGlyAspSerGluGlnGlnLeuValGlu.. 212
  |||||
535 CTGACACAATATCGCT.....GGGACCTTAAGAAACCATTTGTGAGA 578
213 .....AlaPheGluGlnMetThrArgAsnLeuPheSerLeuPro 226
  |||||
579 ATTCAAACTGCATTTCTTCAAGCTTGTGTGGACTTATCCGTTCCGA 628
226 lAspValProPheSerGlyLeuTyrArgGlyMetLysAlaArgAsnLeu 242
  |||||
629 TTGATTTTCCGGGCAAAATTTATCGTTGGGAAATCCACCAAGAAATTAAC 678
243 lIleHisAlaArgIleGluGlnAsnIleArgAlaLysIleCysGlyLeuAr 259
  ::|||
679 ATTGATAGGTGCTTACAGAGCTGATGCAGAAACGT..... 714
259 gAlaSerGluAlaGlyGlnGlyCysLysAspAlaLeuGlnLeuLeuIleG 276
  ::|||
715 ...AGAGATTCTGGAGAAACATTCACAGACATGTAGCTTACTTGATGA 760
276 lHisSerTrpGluArgGlyGluArgLeuAspMetGlnAlaLeuLysGln 292
  ::|||
761 ACAGAGAGAGTAAACCGATACCCGTTAACCGAT...GAAGAGATTAAGAGAC 807
293 SerSerThrGlnLeuLeuPheGlyGlyHisGlnThrThrAlaSerAlaAl 309
  ::|||
808 CAAGTTGTACGATTTTGTATTCGGGTTACGAACGTCTCTACACATC 857
309 aThrSerLeuIleThrTyrLeuGlyLeuTyrProHisValLeuGlnLysV 326
  ::|||
858 AATGATGGCTCTTAAGTACCTTCATGATCAACCCAAAGCTCTTCAAGAAC 907
326 aArgGluGlnLeuLysSerLysGlyLeuLeuCysLysSerAsnGlnAsp 342
  ::|||
908 TAAGAGCTGAG.....CATTTGGCATTCAGGAAACAAACAGACAGGAC 951
343 AsnLysLeuAspMetGlnIleLeuGlnGlnLeuLysTyrIleGlyCysVa 359
  ::|||
952 GAACACATTCGCTCTTGAGAGACGTGAAGTCAATGAAGTTCACTCGAGCTGT 1001
359 lIleLysGlnThrLeuArgLeuAsnProProValProGlyLysPheArgV 376
  |||||
1002 GATTTATAGACATCAAGATGGCAACGATCGTTAATGGGCTCAAGGA 1051
376 aAlaLeuLysThrPheGlyLeuLeuAsnGlyTyrGlnIleProLysGlyTrp 392
  ::|||
1052 AAACATCTGCTGACTTGGAAATCAACGGTTATTATTATCCCAAAAGATGG 1101
393 AsnValIleTyrSerIleCysAspThrHisAspValAlaGluIlePheTh 409
  ::|||
1102 AGAATTTATGTATACACAGGAAATTAATTAACGATGCAATCTTATAGA 1151
409 rAsnLysGlnGluProPheAsnProAspArgPheSerAlaProHisProGluA 426
  ::|||
1152 AGACCATTTGATCTTTAATCATGATGATGAAGAAAGCTTGAG.. 1200
426 sPAlaSerArgPheSerPheIleProPheGlyGlyLeuArgSerCys 442

```

```
||||: ||| :||| ||||| ||| |||
1201 .....TCACAAACCTACCTTGTGTTGGAGTGGACGACGCTTGT 1245
443 VALGlyGluPhaIalysIleuLeuLysIlePhethValGlu 459
|||||: ||| :||| ||||| |||
1246 CCTGGTAAGAACTAGGGATTGTCAGATCTGACCTTTCATTACTT 1295
459 uAlaArgHisCysAspTrpGlnLeuLeuAnGlyProProthMetLysT 476
||||: ||| :||| ||||| |||
1296 TGTAGACATACAGATGGAGCAATAGACGGATGATTAATG.... 1341
476 hSerProthValTyPro 482
1342 .....GTGTTCCG 1350

seq_name: /SIDSL/gcdata/geneseq/geneseqn/NA2000.DAT: AAC42516
seq_documentation_block:
ID AAC42516 standard; DNA; 1648 BP.
XX AAC42516;
AC
XX
XX 17-OCT-2000 (first entry)
DT
XX
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 35853.
DE
XX
XX Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
XX Arabidopsis thaliana.
OS
XX
XX EP1033405-A2.
PN
XX
XX 06-SEP-2000.
PD
XX
XX 25-FEB-2000; 2000EP-0301439.
PF
XX
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
```

```
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140685.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
```

PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149733.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161932.
PR 28-OCT-1999; 99US-0161933.
PR 29-OCT-1999; 99US-0162142.

alignment_scores: Quality: 436.50 Length: 504
 Ratio: 1.480 Gaps: 13
Percent Similarity: 58.532 Percent Identity: 27.381

alignment_block:
US-09-668-482-4 x AAC42516 ..

Align seg 1/1 to: AAC42516 from: 1 to: 1648

5 AlaLeuLeuAlaSerAlaLeuCysThrPheValLeuProLeuLeuPhe 21
||||: ||||: ||| ||| ||| |||
176 GCTATGACGGAACGGGTTTGTATTCATCGTGGTCCCTTGTATATAT 225
21 e.....LeuAlaAlaIleLysLeuTrpAspL 30
 : |||| ||| ||| ||| |||
226 GGGATTGTTGTTTGAATGGGCTGCGAAGAGTGAATGTTTGATT 275
30 eutyrCysValSerGlyArgAspArgSerCysAlaLeuProLeuProp 46
 ||||: ||| ||| ||| ||| |||
276 ATGTGTCCAAACTGGTGGAAAAAACACTAT.....CTGCCACCA 316
47 GlyThrMetGlyPheProPhePheGlyIuThrLeuGlnMetVal..... 61
||| ||| ||| ||| ||| ||| |||
317 GGTGATTGGATGGCCAGTCATAGCACAACATGTGGTCTTCTTGAGC 366
62LeuGlnArgArgIuLysPheLeuGlnMetLysArgArgLysT 75
367 TTTCAAACATCTGATCCGAAATCCCTTCATCCATCCATCACCACGAT 416
75 yrgIy.....PheIleTyLysThrHisLeuPheGlyArgProThrVal 89
|||| |||| |||| |||| |||| ||||
417 ATGGGGGTACCGGATTTATTAAGCACACATGTTGGTACCATTGTGA 466
90 ArgValMetGlyAlaAspAsnValArgArgIleLeuLeuGlyAspAsp 106
 ||| ||| ||| ||| ||| |||
467 CTATGAACAACACACAGACTGTAGCGACTTCAACAGATGATGATC 516
106 gLeuValSerValHisTrpProAlaSerValArgThrIleLeuGlnSerG 123
 :: :: ||| ||| ||| ||| |||
517 CTTC...CACATAGGTGGCCAAATCTACATGAACATCATTTGGAGGA 563
123 LysLeuSerAsnLeuHisAspSerSerHisLysGlnArgLysLysVal 139
 :: :: |||| |||| |||| ||||
564 AGTCCTTGTGTGTAATCCCTTGAAGAACACAGAGGTGAGGGTTTG 613
140 IleMetArgAlaPheSer...ArgIleAlaLeuGlnCysTyrValProVa 155
 :: :: |||| |||| |||| |||
614 ACTCTGCTCTGTCATATGCCCCGTAAGCTCTCTGTATATCATACACT 663
155 IleThrGlnGluValGlySerSerLeuGlnIuThrPheLeuSerCysGly 172
||| ||| |||| |||| |||| |||
664 CATTGAAGAACTGTTAATACGATCTGAAAAATGTCACCAAAATGGGAG 713
172 LuArgGlyLeuLeuValItyrProGluValLysArgLeuMetPheArgIle 188
 || ||| ||| ||| ||| ||| |||
714 AAATCGAGTTCTTATCTCAT.....TTGCCGTAAGCTTAAAGTTC 757
189 AlaMetArgIleLeuLeuGlnCysGluProGlnIleuAlaGlyAspGlyAs 205
 || ||| |||| |||| |||| |||
758 ATTATGTACATATTTCTCGACGAGTGAGAGTCATGATGATGAT..... 802
205 pSerGlnGlnLeuLeuValGluAlaPheGlnGluMetThrArgAsnLeuP 222
 ||| ||| ||| ||| ||| ||| |||
803TCATTGAACGGGAGTATACCAACCTTAACATAGAGTTC 842
222 hSerLeuProIleAspValProPheSerGlyLeuTyrArgGlyMetLys 238
 :: ||| |||| ||| ||| ||| |||
843 GAGCAATGGGATTATATCTCTCGGTTTGCTTATCATAGAGACTTTGAAG 892
239 AlaArgAsnLeuIleHisAlaArgIleGlnGlnAsnIleArgAlaLysI 255
 |||| :: ||| ||| ||| ||| |||
893 GCAGGAAAAAACTTGTAGCTGCCTTT...CAGTCCATAGTACTAACCG 939

PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136782.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0138119.
PR 16-JUN-1999; 99US-0138452.
PR 16-JUN-1999; 99US-0138453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142977.
PR 12-JUL-1999; 99US-0143542.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.

PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151308.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.

```

PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160880.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

```

alignment_scores:
  Quality: 433.50      Length: 486
  Ratio: 1.521         Gaps: 13
  Percent Similarity: 58.642   Percent Identity: 26.543

```

alignment_block:

US-09-668-482-4 x AAC48157 ..

Align seg 1/1 to: AAC48157 from: 1 to: 1646

```

19 LeuLeuPheLeuAlaAlaIleLysLeuTrpAspLeuTyrCysValSerGI 35
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
77 CTCCCTCCCTCTCTCCATCCGCGCGGCTTCCTCTCTACTCCGCG.. 124
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
35 YArgAspArgSerCysAlaLeuProLeuProProGlyThrMetGlyPheP 52
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
125 .CGTACACGTTACCGTCGATGGCTGTGCTCCGGGAGCGTTGGTCTCC 173
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
52 roPhePheGlyLutThrLeuGlnMetVal.....LeuGln 63
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
174 CTCTGATAGAGAGACTTTTACGCTGATCGAGGACTTACAAAACAGAGAAC 223
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
64 ArgAlaGlyLysPheLeuGlnMetLysArgAlaGlyTyrGlyPheIleTyr 80
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
224 CCGTAGCCCTTTCATCGACGAGAGATAGCCCGGTACGGTTCGGTTTCAT 273
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
80 sThrHisLeuPheGlyArgProThrValArgValMetGlyAlaAspAsnV 97
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
274 GACGATCTTTTGGTGAACGACGATTTTTCAGCTGACCGCGAAACGA 323
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
97 aLArgAlaIleLeuLeuGlyAspAspArgLeuValSerValHisTrpPro 113
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
324 ACCGGTTTGTCTTCAGAACGAAAGGAGCTTTTGAAGTGTCTTATCTCT 373
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
114 AlaSerValArgThrIleLeuGlySerGlyCysLeuSerAsnLeuHisAs 130
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
374 GCTTCCATTTTGTAACTTTTGGGAAACACTCTCTCTTCTTAAAGG 423
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
130 pSerSerHisLysGlnArgLysLysValIleMetArgAlaPheSerArgG 147
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
424 TTCCTTGGATAAACGATACGACTCTCTCACCATGAGCTTTGCTAATCTT 473
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
147 IuaIleLeuGlyCysTyrValProVal...IleThrGlnGluValAlaGly 162
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
474 CAATCATTAAGAACCATCTCATGCTGATATTGACCGGTAGTCGGGTTT 523
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
163 SerLeuGlnGlnTrpLeuSerCysGlyLysArgGlyLeuLeuValTyrPr 179
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
524 AATCTTGATTTTGGTCTTCT.....CGTGTCTCTCTATGGA 561
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
179 oGluValLysArgLeuMetPheArgIleAlaMetArgIleLeuLeuGlyC 196
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
562 AGAAGCCAAAGAAAGTAACGTTGACGTAACGGTGAACGATTGATGAGCT 611
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
196 ysgLupProGlnLeuAlaGlyAspGlyAspSerGlnGlnLeuValGlu 212
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
612 TTGATCCA.....GGGGAATGAGAGTGAAGATTAAAGAAA 646

```

```

213 AlaPheGlnGluMetThrArgAsnLeuPheSerLeuProIleAspValPr 229
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
647 GAGTATCTTCTTGTCAATCGAAGGCTTCTCTCTCTCTCTCTCTCCCTC.. 694
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
229 oPheSerGlyLeuTyrArg...GlyMetLysAlaArgAsnLeuIleHisA 245
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
695 .TTCTCCACACACTTACCGCAAGACCATCCAAAGCGCGAGAG..... 733
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
245 IaArgGlyLeuGlnAsnIleArgAlaLysIleCysGlyLeuArg..... 259
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
734 ...AAGTGGCGGAGGCGTTCGCGTGTGATGAAAAGAGAGAGAG 781
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
260 AlaSerGluAlaGlyLysCysLysAspAlaLeuGlnLeuLeuIle... 275
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
782 GAGAGAAAGAGAGCGGAGAGAAAGAAAGATATCTTCGCGCTTCGCTTCG 831
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
276 ...GluHisSerTrpGluArgGlyLysArgLeuAspMetGlnAlaLeu 291
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
832 GCGGAGATGATGATTTTCCGATGAGAGATTTGTGACTTC..... 871
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
291 ysgLnsSerSerThrGluLeuLeuPheGlyGlyHisGlnThrThrAlaSer 307
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
872 .....TTGGTGGCTTACTTGTGCGCGGTTATGAAACACCTCCAG 913
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
308 AlaAlaThrSerLeuIleThrTyrLeuGlyLeuTyrProHisValLeuG 324
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
914 AATGATACCTCGCCGCTCAATTTCTACCGAGACCTCTTACCTCTTCG 963
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
324 nLysValArgGluGlnLeuLysSerLysGlyLeuLeuCysLysSerAsnG 341
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
964 TCACACTCAAGAGAGAG.....CATGAAAAGATTAGGCAATGAAGAGTG 1007
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
341 IAspAsnLysLysAspMetGluIleLeuGlnGlnLeuLysTyrIleGly 357
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1008 ATTGCGATATGCTCTGTAATGAGATGATTACAAGTCATGCCATTACACAA 1057
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
358 CysValIleLysGluThrLeuArgLeuAsnProProValProGlyLysP 374
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1058 TGTGTGTGTTAATGAGAGCGCTACGAGTGCCTAACATCATCGCGGTGTT 1107
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
374 eArgValAlaLeuLysThrPheGluLeuAsnGlyTyrGlnIleProLysG 391
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1108 CAGACGTCGTAATGACGAGATGTGAGATCAAGGTTATTAATAATTCAAAAG 1157
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
391 IYTrpAsnValIleTyrSerIleCysAspThrHisAspValAlaGluIle 407
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1158 GGTGGAAGTATTTTCATGCTTTAGAGCGGTTTACACCCAAACACAC 1207
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
408 PheThrAsnLysGluGluPheAsnProAspArgPheSerAlaProHisPr 424
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1208 TTCAAAGAGATGCTCGCACTTCAACCTTGAAGATGCGAGACGACTCGGT 1257
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
424 oGluAspAlaSerArgPheSerPheIleProPheGlyGlyGlyLeuArgS 441
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1258 AAGCAGAGGCGCTTCTAATGTGTTCACACCGTTGTGTGAGGAGGCAAGGC 1307
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
441 eCysValGlyLysGluPheAlaLysIleLeuLeuLysIlePheThrVal 457
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1308 TATGTCCCGGTTAGAGAGCGGTAGGGTTGCACTCTCTGTCTTCTTCAAC 1357
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
458 GluLeuAlaArgHisCysAspTrpGlnLeuLeuAsnGlyProProThrWe 474
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1358 CGCCTAGTGACAGCGCTCAGTTGG.....GTTCCGTGCAGA 1392
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
474 TlysThrSerProThrValTyrProValAspAsnLeuProAlaArgPheT 491
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1393 GCAAGACAAAGCTGGTTTCTTCCAACTACAAAGAGCAAGAAACGTTACC 1442
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
491 hrHisPhe 493
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1443 CGATCTTC 1450

```

seq_name: /SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT: AAC47986
seq_documentation_block:
ID AAC47986 standard; DNA; 1649 BP.
XX AAC47986;
XX
DT 18-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 55845.
DE
XX Hybridisation assay; genetic mapping; gene expression control;
KM protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX Arabidopsis thaliana.
XX
XX EPI03405-A2.
PN
XX
PD 06-SEP-2000.
XX
XX
XX 25-FEB-2000; 2000EP-0301439.
PF
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-013180.
PR 09-MAR-1999; 99US-0133548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.

PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142035.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144003.
PR 16-JUL-1999; 99US-0144088.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145066.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0148368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.

```
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149920.
PR 23-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159684.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161820.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161933.
PR 29-OCT-1999; 99US-0162142.

alignment_scores:
    Quality: 433.50      Length: 486
    Ratio: 1.521        Gaps: 13
    Percent Similarity: 58.642    Percent Identity: 26.543

alignment_block:
US-09-668-482-4 x AAC47986 ..

Align seg 1/1 to: AAC47986 from: 1 to: 1649
```

```
19 LeuLeuPheLeuAlaIaIeLysLeuTrpAspLeuTyrCysValSerGI 35
|||||:|||||:|||||:|||||
79 CTCCTCTCCCTCTTCATGCGCGCGGCTTCTCTCTCTACTCCGCG. 126
35 YARGSPARGSERCYSLALEUPROLEUPROGLYTHRMETGLYPHEP 52
```

```
127 .GCTACACGTTACCGTCGGATGGCTCTCTCCGGAAGCGTGGCTCC 175
||| |||
52 roPhePheGlyIuThrLeuGlnMetVal.....LeuGln 63
|||||:|||||:|||||:|||||
176 CTCGTATGAGAGACTTTTACAGCTGATCGAGCTTACAAACAGAGAAC 225
64 ARGATGlyPheLeuGlnMetLysArgArgLysTyrGlyPheIleTyrI 80
||| |||:|||||:|||||:|||||
226 CCGAGCCTTTCATCGACGAGAGATGACCGGTACGGTTCGGTTTCAT 275
80 sThrHisLeuPheGlyArgProThrValArgValMetGlyAlaAspAsn 97
|||||:|||||:|||||:|||||
276 GAGCGATCTTTTGGTGAACCGACGATTTTTCAGCTGACCGGAAACGA 325
97 aLArgArgIleLeuLeuGlyAspAspArgLeuValSerValHisTrpPro 113
||| |||:|||||:|||||:|||||
326 ACCGGTTTCTCTCAGAAAGGAGACCTTTTGAGTGTCTTATTCCT 375
114 AlaSerValArgThrIleLeuGlySerGlyCysLeuSerAsnLeuHis 130
|||||:|||||:|||||:|||||
376 GCTTCATTTGTAACTTTTGGGAGAACCTCTCTCTTATGAAAGC 425
426 TTCTTTCATTAACGATGACACTCTCTACCATGAGCTTGTATTCCT 475
147 LuAlaLeuGlnCysTyrValProVal...IleThrGlnIuValGlySer 162
||| |||:|||||:|||||:|||||
476 CATCATTAAGACCATTCATGCTGATTTGACCGGTATTCGGTTT 525
163 SerLeuGlnGlnTrpLeuSerCysGlyIuArgIleLeuValTyrPr 179
||| |||:|||||:|||||:|||||
526 AATCTTGATTTGGTCTCT.....CGTCTCTCTCATGA 563
179 oGluValLysArgLeuMetPheArgIleAlaMetArgIleLeuLeuGlyC 196
|||||:|||||:|||||:|||||
564 AGAAGCCAAAGATPACGTTGAGCTTAACGGTGAAGCATGATGAGCT 613
196 YSGluProGlnLeuAlaGlyAspGlyAspSerGlnGlnIleuValGlu 212
||| |||:|||||:|||||:|||||
614 TTGATCCA.....GGGAATGAGTGAAGCTTAAAGAA 648
213 AlaPheGlnGlnMetThrArgAsnLeuPheSerLeuProIleAspValPr 229
||| |||:|||||:|||||:|||||
649 GAGTATCTTCTTCATCGAAGGCTTCTCTCTCTCTCTCTCTCTCTCT 696
229 oPheSerGlyLeuTyrArg...GlyMetLysAlaArgAsnLeuIleHis 245
|||||:|||||:|||||:|||||
697 .TTCCTCAACACTTACCGCAAGCCATCCAGCGCGAGG..... 735
245 IaArgIleGlnGlnAsnIleArgAlaLysIleCysGlyLeuArg..... 259
||| |||:|||||:|||||:|||||
736 ..AAGGTGCGGAGCGCTTGAACGCTGTGATGATAAAGGAGGAGAG 783
784 GAGGAGAAAGAGCGGAGAGAAAGATATGCTTGGCGGCTTGCCTTGC 833
276 ...GluHisSerTrpGluArgGlyGluArgLeuAspMetGlnAlaLeuL 291
||| |||:|||||:|||||:|||||
834 GCGGATGATGATTTTCCGATGGAAGATTTGCTTCTCT..... 873
291 YSGlnSerSerThrGlnLeuLeuPheGlyGlyHisGlyThrThrAlaSer 307
||| |||:|||||:|||||:|||||
874 .....TTGTGCTTACTTGTCCGCGGCTTATGAACAACCTCCACG 915
308 AlaIaIaThrSerLeuIleThrTyrLeuGlyLeuTyrProHisValLeuG 324
||| |||:|||||:|||||:|||||
916 ATCATGACTCTCGCGGCTCAAAATTTCTACCGAGACTCTTAAAGCTTGC 965
324 nLysValArgGlnIuLeuLysSerLysGlyLeuLeuCysLysSerAang 341
||| |||:|||||:|||||:|||||
```

```
966 TCACCTCAGAGAGAG....CATGAAAGATTAGGCAATGAGAGTG 1009
341 lnaspsnlylseuaspmetgluileugluInleuIysTYrilegly 357
      :::: ::::: ::::: ::::: :::::
1010 ATTGCTATAGTCTTGATAGGATGATTTACAGATCAATGCCATTCACACAA 1059
358 CysValIleIysgluThrleuArgleuAsnProProValProglyIyPh 374
      ||||| ::::: ||||| ::::: ::::: ||| ||
1060 TGTGTGTGTATGAGACGCTACGAGCGGCTAACATCATCGCGGTGTTT 1109
374 eargvAlAlAlleuIysThrPhegluleuAsnglyTYrGlnIleProIysG 391
      |||| |::: |::: |::: |::: |::: |::: |:::
1110 CAGACGTGCATGACGATGTGAGATCAAGSTTATAAATTCCAAG 1159
391 lYTPasValIleTYrSerIleCysaspThrhisaspValAlAluile 407
      ||||| ::::: ||::: ::::: ::::: :::::
1160 GGTGGAAGATATTCTCATGCTTAGAGCGGTTCATTAGACCCAACAC 1209
408 PheThrAsnIysgluIupheAsnProAspArgPheSerAlaProHispr 424
      || |::: ::::: ||||| ||| ::::: :::::
1210 TTCAAAGATGCTCGACTTCAACCTTGAGATGGCAGAGCAACTCGGT 1259
424 oGluaspAlaSerArpPheSerPheIleProPheglyIyglyIleuArgS 441
      :::: ::::: ||| ||||| ||||| |||
1260 AACGACAGGCCCTTCTATGTGTTCAACCGCTTGTGAGGGCCAAAGC 1309
441 erCysValIglyIysgluPheAlAlIysIleleuIysIlePheThrVal 457
      || | || |::: ||::: ::::: ||::: |||
1310 TATGTCCCGGTTACGAGCTGAGGTTGCACCTCTGTTTCTTCAC 1359
458 GluleuAlaArgHisCysaspTrpGlnleuIeAsnglyProProthime 474
      ::::: ::::: ||| |::: |::: |::: |:::
1360 CGCCTAGTGACAGGCTTCAGTTGG.....GTTCCTGCAGA 1394
474 lYsThrSerProThrValTYrProValaspAsnleuProAlaArgPhe 491
      :::: ::::: ||::: ||::: ||::: ||:::
1395 GCAGACAGAGGTGTTCTTCCAACTACAGAAACGAGAAAGGTACC 1444
491 hrHisPhe 493
      |||
1445 CGATCTTC 1452
```


OM of: US-09-668-482-4 to: Issued_Patents_NA.* out_format : pfs
Date: Nov 6, 2001 2:29 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-O/cg2_1/USPTO.spool/US09668482/runat.05112001.13353.10897/app-query.fasta.1.1680
-Db=Issued_Patents_NA -OPMT=fastap -SUFFIX=rni -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blotsum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR=score-pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ent MINLEN=0 -MAXLEN=200000000
-USER=US09668482 -CGCN1_1_104 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPEXT
-WAIT -THREADS=1

Search information block:

Query: US-09-668-482-4
Query length: 497
Database: Issued_Patents_NA.*
Database sequences: 351203
Database length: 11328999
Search time (sec): 149.780000

score_list:

Sequence	Strid	Orig	Zscore	Escore	Len	Documentation	..
/cg2_6/p/odata/2/ina/6A.COMB.seq:US-08-724-466B-5					2611.00	5650.04	0.0
/cg2_6/p/odata/2/ina/6B.COMB.seq:US-08-882-164D-5					2611.00	5650.04	0.0
/cg2_6/p/odata/2/ina/6A.COMB.seq:US-08-882-164D-31					2437.00	5268.72	8.1e-286
/cg2_6/p/odata/2/ina/6B.COMB.seq:US-08-724-466B-3					1738.00	3743.60	7.2e-201
/cg2_6/p/odata/2/ina/6A.COMB.seq:US-08-882-164D-3					1738.00	3743.60	7.2e-201
/cg2_6/p/odata/2/ina/6B.COMB.seq:US-08-882-164D-38					1712.00	3676.12	4.1e-197
/cg2_6/p/odata/2/ina/6A.COMB.seq:US-08-882-164D-36					1365.00	2925.35	2.7e-155
/cg2_6/p/odata/2/ina/6B.COMB.seq:US-08-724-466B-11					614.00	1314.78	1.4e-65
/cg2_6/p/odata/2/ina/6A.COMB.seq:US-08-882-164D-11					433.50	900.96	1.6e-42
/cg2_6/p/odata/2/ina/6B.COMB.seq:US-08-991-677-3					348.50	713.51	4.3e-32
/cg2_6/p/odata/2/ina/6A.COMB.seq:US-09-103-840A-2					344.00	600.54	8.4e-26
/cg2_6/p/odata/2/ina/6B.COMB.seq:US-09-103-840A-1					344.00	600.52	8.5e-26
/cg2_6/p/odata/2/ina/6A.COMB.seq:US-09-103-840A-1					303.00	511.12	8.0e-21
/cg2_6/p/odata/2/ina/6B.COMB.seq:US-09-103-840A-2					303.00	511.12	8.1e-21
/cg2_6/p/odata/2/ina/6A.COMB.seq:US-09-103-840A-2					289.00	584.14	6.9e-25
/cg2_6/p/odata/2/ina/6B.COMB.seq:US-08-948-564-15					288.50	584.11	6.9e-25
/cg2_6/p/odata/2/ina/6A.COMB.seq:US-08-948-564-3					276.00	555.75	2.6e-23
/cg2_6/p/odata/2/ina/6B.COMB.seq:US-08-948-564-1					260.00	531.27	4.1e-22
/cg2_6/p/odata/2/ina/6A.COMB.seq:US-08-313-075A-49					264.50	531.02	4.3e-22
/cg2_6/p/odata/2/ina/6B.COMB.seq:US-08-882-164D-37					252.50	505.82	1.6e-20
/cg2_6/p/odata/2/ina/6A.COMB.seq:US-09-292-768-67					252.50	505.82	1.6e-20
/cg2_6/p/odata/2/ina/6B.COMB.seq:US-09-292-768-69					251.50	503.64	2.1e-20
/cg2_6/p/odata/2/ina/6A.COMB.seq:US-08-881-784-8					251.50	503.64	2.1e-20
/cg2_6/p/odata/2/ina/6B.COMB.seq:US-09-292-768-3					251.00	499.00	3.8e-20
/cg2_6/p/odata/2/ina/6A.COMB.seq:US-09-616-990-63					245.00	499.32	1.3e-19
/cg2_6/p/odata/2/ina/6B.COMB.seq:US-08-457-274A-1					244.50	489.32	1.3e-19
/cg2_6/p/odata/2/ina/6A.COMB.seq:PCR-US95-05758-1					244.50	489.32	1.3e-19
/cg2_6/p/odata/2/ina/6B.COMB.seq:US-09-091-432-3					244.50	473.34	1.0e-18
/cg2_6/p/odata/2/ina/6A.COMB.seq:US-07-912-900-29					242.50	483.31	2.9e-19
/cg2_6/p/odata/2/ina/6B.COMB.seq:US-08-283-309-29					242.50	483.31	2.9e-19
/cg2_6/p/odata/2/ina/6A.COMB.seq:US-08-502-046-29					242.50	483.31	2.9e-19
/cg2_6/p/odata/2/ina/6B.COMB.seq:US-09-033-055A-3					240.00	477.82	5.8e-19
/cg2_6/p/odata/2/ina/6A.COMB.seq:US-09-292-768-63					240.00	477.81	5.8e-19
/cg2_6/p/odata/2/ina/6B.COMB.seq:US-08-457-274A-27					238.50	476.23	7.1e-19
/cg2_6/p/odata/2/ina/6A.COMB.seq:PCR-US95-05758-27					238.50	476.23	7.1e-19
/cg2_6/p/odata/2/ina/6B.COMB.seq:US-08-313-075A-29					238.50	475.29	8.0e-19
/cg2_6/p/odata/2/ina/6A.COMB.seq:US-08-948-564-9					238.50	474.34	9.0e-19
/cg2_6/p/odata/2/ina/6B.COMB.seq:US-08-948-564-17					237.00	471.71	1.3e-18

/cg2_6/p/odata/2/ina/6B.COMB.seq:US-08-991-677-1 + 235.00 467.32 2.2e-18 1708
/cg2_6/p/odata/2/ina/6A.COMB.seq:US-08-881-784-5 + 235.00 466.91 2.3e-18 1762
/cg2_6/p/odata/2/ina/6B.COMB.seq:US-09-292-768-1 + 235.00 466.91 2.3e-18 1762
/cg2_6/p/odata/2/ina/6A.COMB.seq:US-08-948-564-5 + 234.50 466.74 2.4e-18 1644
/cg2_6/p/odata/2/ina/6B.COMB.seq:US-09-292-768-65 + 234.00 464.72 3.1e-18 1762

seq_name: /cg2_6/p/odata/2/ina/6A.COMB.seq:US-08-724-466B-5

seq_documentation_block:

Sequence 5, Application US/08724466B
Patent No. 6063606
GENERAL INFORMATION:
APPLICANT: Petkovich, P. Martin, White, Jay A.,
Beckett, Barbara R., Jones, Glenville
TITLE OF INVENTION: Retinoid Metabolizing Protein
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESSEE: Blake, Cassels & Graydon
STREET: Box 25, Commerce Court West
CITY: Toronto
ZIP: M5L 1A9
COUNTRY: Canada

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 Mb storage
COMPUTER: COMPAQ, IBM PC compatible
OPERATING SYSTEM: MS-DOS 5.1

SOFTWARE: WORD PERFECT

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/724,466B

FILING DATE: October 1, 1996

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/667,546

FILING DATE: June 21, 1996

ATTORNEY/AGENT INFORMATION:

NAME: Hunt, John C.

REGISTRATION NUMBER: 36,424

REFERENCE/DOCKET NUMBER: 50767/00004

TELECOMMUNICATION INFORMATION:

TELEPHONE: (416) 863-2653

TELEFAX: (416) 863-2653

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 1494 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-724-466B-5

alignment_scores:

Quality: 2611.00

Ratio: 5.254

Percent Similarity: 100.000

Percent Identity: 100.000

Length: 497

Gaps: 0

Align seg 1/1 to: US-08-724-466B-5 from: 1 to: 1494

1 MetcylleuProAlaIleuLeuAlaSerAlaLeuCyThrPhaIleuR 17

1 ATGGGCTCCCGCGCTGCGCCAGCGCCCTGACCTTCGTGCTGCC 50

17 OleuleuLeuPheLeuAlaIleuLeuLeuTrpAspIleuTrpCysValS 34

1 GCTGCTCTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 100

51 GCTGCTCTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 100

34 ergIyArGAsPArGSeCyAlaLeuProLeuProGlyThrMetGly 50

101 GCGGCGCGACGACGATGTCCTCCATTCGCCCGCGGACTATGAGGC 150

51 PheProPhePheGlyGlyTrpLeuGlnMetValLeuGlnArgTrgysph 67

1 PheProPhePheGlyGlyTrpLeuGlnMetValLeuGlnArgTrgysph 67

```

151 TTCCCTTCCTTTGGGAAACCTTCAGATGATCTGCACGGGAGAACTT 200
67 eLeuGlMetLysArgArgLysTyrGlyPheIleTyrLysThrHisLeuP 84
|||||
201 CCTGCAGATGAAGCCGAGAAATACGGCTTCTCATACAGACGCACTCTG 250
84 heGlyArgProThrValArgValMetGlyAlaAspAsnValArgArgLe 100
|||||
251 TCGGGCGGCCACCGTACGGGTGATGGCGGACAAATGTGGCGGCATC 300
101 LeuLeuGlyAspAspArgLeuValSerValHisThrProAlaSerValAr 117
|||||
117 gThrIleLeuGlySerGlyCysLeuSerAsnLeuHisAspSerSerHisL 134
|||||
351 CACCATTTCTGGGATCTGGGTGCTCTCTACCTGCACGACCTCTCTGCA 400
134 ysgLArgLysLysValIleMetArgAlaPheSerArgGluAlaLeuGlu 150
|||||
401 AGCAGCGCAAGAGGTGATTATGCGGGCTTCAGCCGCGAGGCACTCGAA 450
151 CysTyrValProValIleThrGluGluValGlySerSerLeuGluTr 167
|||||
451 TGCTACGTGCGGTGATCCGAGGAAGTGGCGACAGCTGAGAGCAATG 500
167 pLeuSerCysGlyLysArgGlyLeuLeuValTyrProGluValLysArgL 184
|||||
501 GCTGAGCTGCGCGGAGCGGGCTCTGCTGCTACCCCGAGGTGAAGCGCG 550
184 eumePheArgGlyLeuAlaMetArgIleLeuLeuGlyCysGluProGluLeu 200
|||||
551 TCATGTCCGAAATCCCATGCGCATCTCACTGGGCTGCGAAACCCAACTG 600
201 AlaGlyAspGlyAspSerGluGluGluLeuValGluAlaPheGluGlu 217
|||||
601 GCGGGCGAGCGGGAGCTCCGAGCAGCACTTGAGAGCCCTTGAGAGAAAT 650
217 tThrArgAsnLeuPheSerLeuProIleAspValProPheSerGlyLeu 234
|||||
651 GACCGCGCATCTCTCTGCTGCTGCCATCGACGTGCCCTTACCGGGCTGT 700
234 yArgGlyMetLysAlaArgAsnLeuIleHisAlaArgIleGluGluAsn 250
|||||
701 ACCGGGGATGAAGCGCGGAACTTCATTCACGCGGATCAGAGAAC 750
251 IleArgAlaLysIleCysGlyLeuArgAlaSerGluAlaGlyGluGly 267
|||||
751 ATTCGCGCGCAAGATCTGGGGCTGCGGGCATCCGAGCGGCGCAGGCTG 800
267 sLyAspAlaLeuGluLeuLeuIleGluHisSerTrpGluArgGlyGlu 284
|||||
801 CAAAGACCGCTGCGAGCTGTGATCGACACTCGTGGGAGAGGAGAGC 850
284 rGleuAspMetGlnAlaLeuLysGlnSerSerThrGluLeuPheGly 300
|||||
851 GCGTGGAGATCGAGCACTAAAGCAATTTCAACCGAACTCTCTTTTGA 900
301 GlyHisGluThrThrAlaSerAlaAlaThrSerLeuIleThrTyrLeuG 317
|||||
901 GGACACGAAACACGCGCATGTCAGACCATCTCGATCATTACTTACTGG 950
317 yLeuTyrProHisValLeuGluLysValArgGluGluLeuLysSerLys 334
|||||
951 GCTTACCCCATGTCTTCAGAAAGTGCAGAGAGCTGAAGAGTAAGG 1000
334 yLeuLeuCysLysSerAsnGlnAspAsnLysLeuAspMetGluLeuLeu 350
|||||
1001 GTTACTTCTTGAAGAGCAATCAAGCAAAAGTTGGACATGGAATTTTG 1050
351 GluGluLeuLysTyrIleGlyCysValIleLysGluThrIleArgLeuAs 367
|||||
1051 GAACAACTTAATATCATCGGGTGTATTATTAAGAGAGACCTTCGACTGA 1100

```

```

367 nProProValProGlyGlyPheArgValAlaLeuLysThrPheGluLeu 384
|||||
1101 TCCCCAGATTCAGAGAGGTTTGGGTGTCTGAAGACTTTGGAATTAA 1150
384 snGlyTyrGlnIleProLysGlyTyrPAsnValIleTyrSerIleCysAsp 400
|||||
1151 ARGATACCAAGATTCACCAAGGCGTGAATGTATCTACAGTATCTGTGAT 1200
401 ThrHisAspValAlaGluIlePheThrAsnLysGluGluPheAsnProAs 417
|||||
1201 ACTCATGATGTGGCAGAGATCTTCACCAACAGGAAGAAATTAATCTGA 1250
417 pArgPheSerAlaProHisProGluAspAlaSerArgPheSerPheIleP 434
|||||
1251 CCGATTCACTGCTCTCCACCCAGAGATGATCCAGAGTTCAAGTTATTC 1300
434 ropheGlyGlyLeuArgSerCysValGlyLysGluPheAlaLysIle 450
|||||
1301 CATTTGAGAGAGCGCTTAGAGCTGTGTAGCCAAAGAAATTTGCAAAAT 1350
451 LeuLeuLysIlePheThrValGluLeuAlaArgHisCysAspTrpGluLe 467
|||||
1351 CTTCTCAAAATATTACAGTGAAGTGGCGCAGCATTTGACTGGCAGCT 1400
467 uLeuAsnGlyProProThrMetLysThrSerProThrValTyrProValA 484
|||||
1401 TCTAATGAGACTCTCTACAAATGAAGAAACAGTCCACACCGTATCTGTG 1450
484 sPAsnLeuProAlaArgPheThrHisPheHisGlyGluLe 497
|||||
1451 ACAATCTCCCTGCAAGATTCACCCATTTCCATGGGGGAATC 1491
seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-882-164D-5
seq_documentation_block:
: Sequence 5, Application US/08882164D
: Patent No. 6306624
: GENERAL INFORMATION:
: APPLICANT: Petkovich, P. Martin, White, Jay A.,
: APPLICANT: Beckett, Barbara R., Jones, Glenville
: TITLE OF INVENTION: Retinoid Metabolizing Protein
: NUMBER OF SEQUENCES: 43
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Blake, Cassels & Graydon
: STREET: Box 25, Commerce Court West
: CITY: Toronto
: STATE: Ontario
: COUNTRY: Canada
: ZIP: M5L 1A9
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 Mb storage
: COMPUTER: COMPAQ, IBM PC compatible
: OPERATING SYSTEM: MS-DOS 5.1
: SOFTWARE: WORD PERFECT
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/882,164D
: FILING DATE: June 25, 1997
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/667,546
: FILING DATE: June 21, 1996
: APPLICATION NUMBER: 08/724,466
: FILING DATE: October 1, 1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Hunt, John C.
: REGISTRATION NUMBER: 36,424
: REFERENCE/DOCKET NUMBER: 50767/00010
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (416) 863-4344
: TELEFAX: (416) 863-2653
: INFORMATION FOR SEQ ID NO: 5:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1494 base pairs

```

```

;
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-882-164D-5

```

```

alignment_scores:
    quality: 2611.00      Length: 497
    Ratio: 5.254          Gaps: 0
Percent Similarity: 100.000    Percent Identity: 100.000

```

```

alignment_block:
US-09-668-482-4 x US-08-882-164D-5 ..

```

```

Align seg 1/1 to: US-08-882-164D-5 from: 1 to: 1494

```

```

1 MetGlyLeuProAlaLeuAlaSerAlaLeuCysThrPheValLeuPr 17
1 ATGGGGCTCCGGGCTGCTGGCCAGTGGCTTGCACCTTCGTGCTGCC 50
17 OleuLeuPheLeuAlaAlaIleLysLeuTrpAspLeuTyrCysValS 34
51 GCTGCTCTCTTCTGCTGGCTGCATCAAGCTCTGGACCTGTACTGCTGA 100
34 erGlyArgAspArgSerCysAlaLeuProLeuProGlyThrMetGly 50
101 GCGGCCGCGACCGCAGTTGTGCTCCCATTCGCCCGGCGACTATGGGC 150
51 PheProPhePheGlyGlyThrLeuGlnMetValLeuGlnArgArgLys 67
151 TTCCTCTCTTGTGGGAACCTTGCAAGATGTACTGACGCGGAGGAAGTT 200
67 eLeuGlnMetLysArgArgLysTyrGlyPheIleTyrLysThrHisLeu 84
201 CCTGCAGATGAAAGCGAGAAATACGGCTTCATCTACAAAGACCATCTGT 250
84 heGlyArgProThrValArgValMetGlyAlaAspAsnValArgArgL 100
251 TCGGGCGGCCACCGTACGGGTATGGGCGGAGCAATGTGGCGGCGATC 300
101 LeuLeuGlyAspAspArgLeuValSerValHisTrpProAlaSerVal 117
301 TTGCTCGGAAGACGACCGGCTGTGCTGCTGCACAGCGCAGCTGGTGGC 350
117 gThrIleLeuGlySerGlyCysLeuSerAsnLeuHisAspSerSerHis 134
351 CACCATCTGGGATCTGGCTGCTCTTAACCTGCACAGCTCTCCGCA 400
134 ysgInArgLysLysValIleMetArgAlaPheSerArgGluAlaLeuG 150
401 AGCAGCGCAAGAGGTATATGCGGGCTTCACCGCGAGGCACTGGAA 450
151 CysTyrValProValIleThrGlnGluValGlySerSerLeuGlnTr 167
451 TGTACAGCGCGGCTGATCACCGAGAAAGTGGGCGAGCGCTGGAGCAGTG 500
167 pLeuSerCysGlyLysArgGlyLeuLeuValTyrProGluValLysArg 184
501 GCTAGAGTGGCGGAGCGCGGCTCTGCTTACCCGAGGTGAAGGCC 550
184 eumetPheArgIleAlaMetArgIleLeuLeuGlyCysGluProGlnLeu 200
551 TCATGTTCCGAATGCCATGCGCATCTCACTGGGCTGGGAACCCCAACTG 600
201 AlaGlyAspGlyAspSerGluGlnGlnLeuValGluAlaPheGlnGlu 217
601 GCGGCGAGCGGAGACTCGAGCAGACAGCTGTGGAGGCTTCGAGGAAT 650
217 tThrArgAsnLeuPheSerLeuProIleAspValProPheSerGlyLeu 234
651 GACCGGCAATCTTTCTGCTGCTGCCATGACGCTTCCTTCAGCGGGCTGT 700
234 yTArgGlyMetLysAlaArgAsnLeuIleHisAlaArgIleGluGlnAsn 250

```

```

|||||
701 ACCGGGCGATAGCGCGGAAACCTCATTCACGCCGATGACAGACAAC 750
251 lIearGalAlaLysIleCysGlyLeuArgAlaSerGluAlaGlyGlnGly 267
|||||
751 ATTGCGCCCAAGATCTGCGGGCTGGGCGATCCGAGGGCGGCGGCTG 800
267 sLysAspAlaLeuGlnLeuIleGlnHisSerTrpGluArgGlyGlu 284
801 CAAGACGCGCTGCGAGCTGTGATGACACTCTGTGGAGAGGGGAGAGC 850
284 rGLeuAspMetGlnAlaLeuLysGlnSerSerThrGluLeuPheGly 300
851 GCGTGCATGACGAGGCACTAAAGCAATCTCAACGCACTCCTTTTGA 900
301 GlyHisGlyThrThrAlaSerAlaAlaThrSerLeuIleThrTyrLeu 317
901 GGACACGAAACCCAGCCGACGACCCACATCTGTGATCATTACTTGG 950
317 yLeuTyrProHisValLeuGlnLysValArgGluLeuLysSerLysG 334
951 GCTTACCCACATGTTCTCCAGAAAGTCCGAGAACAGCTGAAGAGTAAG 1000
334 lYLeuLeuCysLysSerAsnGlnAspAsnLysLeuAspMetGluLeu 350
1001 GTTACTTTGCAAGACAAATCAAGACAACTTGACATGGAATTTTG 1050
351 gLuglnLeuLysTyrIleGlyCysValIleLysGluThrLeuArgLeu 367
1051 GAACAACCTTAATACATCGGTGTGTATTAAAGAGACCTTCGACTGA 1100
367 nProProValProGlyGlyPheArgValAlaLeuLysThrPheGluLeu 384
1101 TCCCCAGTTCAGAGAGGCTTCCGGTCTCTGAAAGCTTTGAATTAA 1150
384 sngLYTyrGlnIleProLysGlyTrpAsnValIleTyrSerIleCysAsp 400
1151 ATGCATACCAAGATTCACAGGCGCTGGAATGTATCTACAGTATCTGAT 1200
401 ThrHisAspValAlaGluIlePheThrAsnLysGluGluPheAsnPro 417
1201 ACTCATGATGTGGCAGAGATCTTCACCAACAAGAGAAATTTAACTCT 1250
417 pArgPheSerAlaProHisProGluAspAlaSerArgPheSerPheIle 434
1251 CCATTCAGTGCCTCTCACCCAGAGATGATCCAGTCCAGCTTCATTC 1300
434 robPheGlyGlyLeuArgSerCysValGlyLysGluPheAlaLysIle 450
1301 CATTTGAGAGAGGCTTAGAGCTGTGTAGGCAAAAGAAATTTGAAAAAT 1350
451 LeuLeuLysIlePheThrValGluLeuAlaArgHisCysAspTrpGln 467
1351 CTTCCTAAAAATATTACAGTGAAGTGGCCAGGCAATTGTGACTGGAGCT 1400
467 uLeuAsnGlyProProThrMetLysThrSerProThrValTyrProVal 484
1401 TCATAATGAGACCTCTACATGAATAACAGATCCACCGTGTATCCGTGG 1450
484 spAsnLeuProAlaArgPheThrHisPheHisGlyGluIle 497
1451 ACAATCTCCCTGCAGATTCACCACTTTCATGAGGGAATC 1491

```

```

seq_name: /cgn2_6/Plodata/2/lna/6B_COMB.seq:US-08-882-164D-31

```

```

seq_documentation_block:
; Sequence 31, Application US/08882164D
; Patent No. 6306624
; GENERAL INFORMATION:
; APPLICANT: Petkovich, P. Martin, White, Jay A.
; APPLICANT: Beckett, Barbara R., Jones, Glenville
; TITLE OF INVENTION: Retinoid Metabolizing Protein
; NUMBER OF SEQUENCES: 43

```

```

CORRESPONDENCE ADDRESS:
ADDRESSEE: Blake, Cassels & Graydon
STREET: Box 25, Commerce Court West
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5L 1A9
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 Mb storage
COMPUTER: COMPAQ, IBM PC compatible
OPERATING SYSTEM: MS-DOS 5.1
SOFTWARE: WORD PERFECT
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/882,164D
FILING DATE: June 25, 1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/667,546
FILING DATE: June 21, 1996
APPLICATION NUMBER: 08/724,466
FILING DATE: October 1, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Hunt, John C.
REGISTRATION NUMBER: 36,424
REFERENCE/DOCKET NUMBER: 50767/00010
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 863-4344
TELEFAX: (416) 863-2653
INFORMATION FOR SEO ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 1725 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-882-164D-31

alignment_scores:
Quality: 2437.00      Length: 497
Ratio: 5.035          Gaps: 0
Percent Similarity: 97.384      Percent Identity: 93.360

alignment_block:
US-09-668-482-4 x US-08-882-164D-31 ..

Align seg 1/1 to: US-08-882-164D-31 from: 1 to: 1725

1 MetGlyLeuProAlaLeuAlaSerAlaLeuCysThrPheValLeuPr 17
|||||
25 ATGGGGCTCCGGCGCTGCTGCGCCAGTGCCTGCTGCTGCTGCTGCC 74
17 OLeuLeuPheLeuAlaAlaLeuLeuTrpAspLeuTyrCysValS 34
|||||
75 GCTGCTGCTCTTCTCGGGCGGCTCAAGCTCTGGGACCTGTACTGTGA 124
34 ergLyArgAspArgSerCysAlaLeuProLeuProProGlyThrMetGly 50
|||||
125 GCAGCGCGAGTCGAGCTGCGCCCTCCCTTGCCTCCCGCGTAACCATGGGC 174
51 PheProPhePheGlyGluThrLeuGlnMetValLeuGlnArgArgLysPhe 67
|||||
175 TTCCCATTTCTTGGGGAACATGTCAGATGTGCTTCAGCGAGGAATT 224
67 eLeuGlnMetLysArgArgLysTyrGlyPheIleTyrLysThrHisLeuP 84
|||||
84 heGlyArgProThrValArgValMetGlyAlaAspAsnValArgArgLys 100
|||||
275 TTGGGCGGCGCCAGCTGGGTGATGGCGCGGATTAATGTCCGGCGCATC 324
101 LeuLeuGlnLysAspArgLeuValSerValHisTrpProAlaSerValAr 117
|||||
325 TTGCTGGAGAGCACCGGTTGTGTGCTGTCACATGCGCCGCGGTGCTGCG 374
```

```

117 gThrIleLeuGlySerGlyCysLeuSerAsnLeuHisAspSerHisL 134
|||||
375 CACATCTCTGGGGCTGCTGCTCCACCTGCAGATTCCTGCAACA 424
134 ySgInArgLysLysValIleMetArgAlaPheSerArgLysAlaLeuGln 150
|||||
425 ACAGCGAAGAAAGTGATTTATGACAGCCTTCAGCCGCGAGGACCTCCAG 474
151 CysTyrValProValIleThrGluGluValAlaGlySerSerLeuGlnTr 167
|||||
475 TCCTACGTCCTGCTGATGCTGAGGAAGTACACAGTTGCTGAGAGCATG 524
167 pleuSerCysGlyGluArgGlyLeuLeuValTyrProGluValLysArgL 184
|||||
525 GCTAAGCTCGCGGAGCGGCGGCTCCTGCTCAACCCGAGGTGAACGCC 574
184 eumetPheArgIleAlaMetArgIleLeuLeuGlyCysGluProGlnLeu 200
|||||
575 TCATGTTCCGATCCGCGCATGCCATCCTGCTGGGCTGCGAGCCGGTCCA 624
201 AlagLyAspGlyAspSerGluGlnGlnLeuValGluAlaPheGluGlu 217
|||||
625 GCGGGCGCGGGGAGAGCGACAGCATCTGTGAGGCTTTGAGAGAT 674
217 tThrArgAsnLeuPheSerLeuProIleAspValProPheSerGlyLeu 234
|||||
675 GACCCGCAATCTCTCTCTCTCCATTGACGTGCGCTTACGGGCTGT 724
234 yTrArgGlyMetLysAlaArgAsnLeuIleHisAlaArgIleGluGlnAs 250
|||||
725 ACCGGGCGCTGAAGCGCGCAACTTATACGCGCGCATCAGAGAGAAC 774
251 IleArgAlaLysIleCysGlyLeuArgAlaSerGluAlaGlyGlnGlyC 267
|||||
775 ATTCGCGCCAGATCCCGCGCTTACGCTACAGACCGGATGGGGTGG 824
267 sLysAspAlaLeuGlnLeuLeuIleGluHisSerTrpGluArgGlyGlu 284
|||||
825 CAAGGCGGCGCTGCAAGCTCTGATTGAGCATCGTGGGAGAGGGAGAGA 874
284 rGLeuAspMetGlnAlaLeuLysGlnSerSerThrGluLeuLeuPheGly 300
|||||
875 GCGTGATATGCGAGCCTTAAACATGCTCAACAGAGAGCTCTCTTGTGT 924
301 GlyHisGluThrThrAlaSerAlaIleThrSerLeuIleThrTyrLeuG 317
|||||
925 GGTCAATGAACATACAGCAATGCTCGACATCATGATCATCTTACCTAG 974
317 yLeuTyrProHisValLeuGlnLysValArgGluGluLeuLysSerLysG 334
|||||
975 ACTTATACCAATGTCCTCCAGAAAGTTCGACAGAGATTAAGACCAAGG 1024
334 yLeuLeuCysLysSerAsnGlnAspAsnLysLeuAspMetGluIleLeu 350
|||||
1025 GCTTACTTTGCAAGAGCAATCAAGCAACAAGTTAGATGAGAACTTTG 1074
351 GluGlnLeuLysTyrIleGlyCysValIleLysGluThrLeuArgLys 367
|||||
1075 GAACAGCTTAAATACATTTGGGTGTCTTAAGAGACCTTCGATTGAA 1124
1125 TCCTCGGCTCCAGAGAGGTTTGGGTGTCTGTAAGACTTTTGGAGTGA 1174
367 nProProValProGlyGlyPheArgValAlaLeuLysThrPheGluLeu 384
|||||
384 snGlyTyrGlnIleProLysGlyTrpAsnValIleTyrSerIleCysAsp 400
|||||
1175 ATGGATACCAATCCCAAGGCTGGAATGTTATTACAGATCTGTGAC 1224
401 ThrHisAspValAlaGluIlePheThrAsnLysGluGluPheAsnProAs 417
|||||
1225 ACCCAGATGTGGCAGATATCTTCACTACAAAGAGGAATTAATCCCGCA 1274
```

```

417 pArgpSerAlaProHisProGluAspAlaSerArgpSerPheIleP 434
||||| :|||||
1275 CCCTTTAATAGTCCATCATCAGAGATGCTTCCCGTTACAGCTTCATTC 1324
434 rohehLygLygLeuArgSerCysValGlyLysGluPheAlaLysIle 450
||||| :|||||
1325 CATTTGAGAGAGCGCTTGAGCTGTGTAGGCAAGAGTTTGCAAAATTT 1374
451 LeuLeuLysIlePheThrValGluLeuAlaArgHisCysAspTrpIle 467
||||| :|||||
1375 CTCTTAAGATATTATACAGAGGAGCTGGTACGACACTGTGATGGACGCT 1424
467 uLeuAsnGlyProProThrMetLysThrSerProThrValTyProValA 484
||||| :|||||
1425 TCTAATGGACCTCTACATGAAGAACAGCCCACTGTGTACCTGTGG 1474
484 spAsnLeuProAlaArgpHeThrHisPheHisGlyLysIle 497
||||| :|||||
1475 ACAATCTCCCTGCAGATTCACCTACTTCCAGGAGATATTC 1515

```

seq_name: /cgn2_6/plodata/2/ina/6A_COMB.seq:US-08-724-466B-3

seq_documentation_block:

```

; Sequence 3, Application US/08724466B
; Patent No. 6063606
; GENERAL INFORMATION:
; APPLICANT: Petkovich, P. Martin, White, Jay A.,
; APPLICANT: Beckett, Barbara R., Jones, Glenville
; TITLE OF INVENTION: Retinoid Metabolizing Protein
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Blake, Cassels & Graydon
; STREET: Box 25, Commerce Court West
; CITY: Toronto
; ZIP: M5L 1A9
; COUNTRY: Canada
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 MB storage
; COMPUTER: COMPAQ, IBM PC compatible
; OPERATING SYSTEM: MS-DOS 5.1
; SOFTWARE: WORD PERFECT
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/724,466B
; FILING DATE: October 1, 1996
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 08/667,546
; FILING DATE: June 21, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Hunt, John C.
; REGISTRATION NUMBER: 36,424
; REFERENCE/DOCKET NUMBER: 50767/00004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 863-4344
; TELEFAX: (416) 863-2653
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1850 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-724-466B-3

```

alignment_scores:

```

Quality: 1738.00      Length: 497
Ratio: 4.080          Gaps: 5
Percent Similarity: 85.714      Percent Identity: 68.008

```

alignment_block:

US-09-668-482-4 x US-08-724-466B-3 ..

Align seg 1/1 to: US-08-724-466B-3 from: 1 to: 1850

```

1 MetGlyLeuProAlaLeuLeuAlaSerAlaLeuCysThrPheValLeuPr 17
||||| :|||||
88 ATGGGGCTGTACACCTTATGCTACCTTTCTTGCACCATCGTGTAC 137
17 OleuLeuLeuPheLeuAlaAlaIleLysLeuTrpAspLeuTrpCysValS 34
||||| :|||||
138 CGTTTACTCTTTCTCGCGCGGTGAAGTTGTGGAGATGTTAATGATCC 187
34 erGlyArgAspArgSerCysAlaLeuProLeuProGlyLysMetGly 50
||||| :|||||
188 GACGAGTGCATCCGAACTGCAGAACTCCTTACCGCCAGGTACCATGGGC 237
51 PheProPhePheGlyGluThrLeuGlnMetValLeuGlnArgLysPhe 67
||||| :|||||
238 TTGCCGTTTCATTGGAGAAAGCGTCCAGCTGATCCTCCAGAGAGAAATT 287
67 eLeuGlnMetLysArgArgLysTyGlyPheIleTrpLysThrHisLeuP 84
||||| :|||||
288 TCTGGCATGAAAGCGAGAAATACGGGTGCATCTCAAGACGCACTCT 337
84 heGlyArgProThrValArgValMetGlyAlaAspAsnValArgArgIle 100
||||| :|||||
338 TCGGAAACCGACTGTACAGGTGATGGAGCTGATATGTGTAGCGCATTT 387
101 LeuLeuGlyAspAspArgLeuValSerValHisTrpProAlaSerValAr 117
||||| :|||||
388 CTGCTGGCGCAACACAGCTGTCTGTCTGTCAGTGGCCAGCATGAGTAG 437
117 gThrIleLeuGlySerGlyCysLeuSerAsnLeuHisAspSerSerHisL 134
||||| :|||||
438 AACCATCTGGGCTGTGACACCTCTCAATGTCATGAGTGAACAACA 487
134 yGlnIArgLysLysValIleMetArgAlaPheSerArgLysLysIleGlu 150
||||| :|||||
488 AAAACAAGAAAAAGCCATTATGAGGCGCTTCTCTGAGATGCTGGAG 537
151 CysTyValProValIleThrGlnGluValGlySerSerLeuGlnIleTr 167
||||| :|||||
538 CACTACATTCCTCGTATCCAGCAGCAGAGGTGAAGGCCCATACAGGAATG 587
167 pleu.....SerCysGlyLysArgGlyLeuLeuValTyProGluVal 181
||||| :|||||
588 GCTGCAAAAAAGACTCTGC.....GTGCTGTGTATCCAGAAA 625
181 allysArgLeuMetPheArgIleAlaMetArgIleLeuLeuGlyCysGlu 197
||||| :|||||
626 TGAAGAAACTCATGTTTGGATAGCTATGAGATCTGCTTGGTTTGA 675
198 Pro...GlnLeuAlaGlyAspGlyAspSerGluGlnGlnLeuValGluAl 213
||||| :|||||
676 CCAAGACCAATTAAGACGAGC.....GAGCAAGAACTGGTGAAC 716
213 apheGlnGluMetThrArgAsnLeuPheSerLeuProIleAspValProp 230
||||| :|||||
717 TTTTGAAGAAATGATCAAAAACTTGTCTCTGCCAATCGAGCTGCTT 766
230 heSerGlyLeuTyArgGlyMetLysAlaArgAsnLeuIleHisAlaArg 246
||||| :|||||
767 TCAAGTGTCTGTACAGGGGTTTGAAGGCAAGCAATTCATCTCTCAAA 816
247 IleGlnGlnAsnIleArgAlaLysIleCysGlyLeuArgAlaSerGluAl 263
||||| :|||||
817 ATTGAGAGAAACATCAAGAGAAATTCAGAT...GAGCAAGAAAGAAA 863
263 agLyGlnGlyCysLysAspAlaLeuGlnLeuLeuIleGlnHisSerTrpS 280
||||| :|||||
864 CGAACAGAAATACAAAGACCCCTTCAGCTGTGATCGAAGACGAGAGA 913
280 LuArgGlyLysArgLeuAspMetGlnAlaLeuLysGlnSerSerTrpGlu 296
||||| :|||||
914 GAAGTACGAACTTTTATGTTTGCAGGCGATGAAGAGACGCTACAGAG 963
297 LeuLeuPheGlyGlyHisGluThrAlaSerAlaAlaThrSerLeuIle 313

```

```

|||||
964 CTTCTATTGGAGGTCATGAACCAACCGCCAGCACTGCACACTCACTTCT 1013
313 eThrTyrlEuGlYLeuTyrrProHisValLeuGlnLysValArgGluLul 330
1014 CATGTTTCTGGCTCTGAACACACAAAGTGTGTCAAGAGGTCAAGAGAGGAG 1063
330 euLysSerTySglYLeuCySlysSerAsnGlnAspAsnLysLysAsp 346
1064 TTCAGAGAAAGGTTGAATGGCATGTATACACCTCGAAAGGGCTTGAGT 1113
317 MetGluLeuGluGlnLeuLysTyrrIleGlyCysValIleLysGluTh 363
1114 ATGAGAGCTGTGGACCACTGTAAGTACACTGATGTGTATTAAAGAGAC 1163
363 rLeuArgLeuAsnProProValProGlyGlyPheArgValAlaLeuLysT 380
1164 TCTTAGAAATCAACCTCTGTTCCGAGGAGATTCAGAGTCGACCTCAAA 1213
380 hrPheGluLeuAsnGlyTyrrGlnIleProLysGlyTrpAsnValIleTy 396
1214 CTTTGAATTTGATGTATCCAAATTTCTAAAGATGGAAGCTCATTTAC 1263
397 SerIleCysAspThrHisAspValAlaGluIlePheThrAsnLysGluL 413
1264 AGCATCTGTGACACGACAGATGTGCGACGCTTTCCAAACAAAGAGAG 1313
413 uPheAsnProAspArgPheSerAlaProHisProGlnAspAlaSerArgP 430
1314 GTTCAGCGCGAGAGATTCATGAGCAAAAGCTGTGAGAGAGCGGTCAGGT 1363
430 hSerPheIleProPheGlyGlyLysLeuArgSerCysValGlyLysGlu 446
1364 TTACTACATCTCCCTTCGAGAGAGATTCAGAGATGTGTGTGGCAAGAG 1413
447 PheAlaLysIleLeuLeuLysIlePheThrValGluLeuAlaArgHisCy 463
1414 TTCGCCAAAGGTTTACTCAAGATCTTTTGTAGTTAAACGACACCATTTG 1463
463 sAspTrpGlnLeuLeuAsnGlyProProThrMetLysThrSerProThrV 480
1464 CAATGTGATCTCTCAACGAGACCCCGACATGAAACAGCCCGCAGCTA 1513
480 aLlyrProValAspAsnLeuProAlaArgPheThrHisPhe 493
1514 TTACCCAGTGAGACATCTCCCTACCAATTCACATGATTAT 1554

seq_name: /cgn2_6/prodata/2/lna/6h_COMB.seq:us-08-882-164D-3
seq_documentation_block:
: Sequence 3, Application US/08882164D
: Patent No. 6306624
: GENERAL INFORMATION:
: APPLICANT: Petkovich, P. Martin, White, Jay A.,
: APPLICANT: Beckett, Barbara R., Jones, Glenville
: TITLE OF INVENTION: Retinoid Metabolizing Protein
: NUMBER OF SEQUENCES: 43
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Blake, Cassels & Graydon
: STREET: Box 25, Commerce Court West
: CITY: Toronto
: STATE: Ontario
: COUNTRY: Canada
: ZIP: M5L 1A9
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 Mb storage
: OPERATING SYSTEM: MS-DOS 5.1
: SOFTWARE: WORD PERFECT
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/882,164D
: FILING DATE: June 25, 1997
: PRIOR APPLICATION DATA:

```

```

: APPLICATION NUMBER: 08/667,546
: FILING DATE: June 21, 1996
: APPLICATION NUMBER: 08/724,466
: FILING DATE: October 1, 1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Hunt, John C.
: REGISTRATION NUMBER: 36,424
: REFERENCE/DOCKET NUMBER: 50767/00010
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (416) 863-4344
: TELEFAX: (416) 863-2653
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1850 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: US-08-882-164D-3

alignment_scores:
Quality: 1738.00 Length: 497
Ratio: 4.080 Gaps: 5
Percent Similarity: 85.714 Percent Identity: 68.008

alignment_block:
US-09-668-482-4 x US-08-882-164D-3 ..
Align seg 1/1 to: US-08-882-164D-3 from: 1 to: 1850

1 MetGlyLeuProAlaLeuLeuAlaSerAlaLeuCysThrPheValLeuPr 17
88 ATGGGGCTGTACACCCCTTATGATGTACCTTCTGTGACCATGCTCTACC 137
17 OleuLeuPheLeuAlaIleLysLeuTrpAspLeuTyrrCysValS 34
138 GCTTTACTCTTTCGCGCGGTGAAGTTGTGGAGATGTTAATGATCC 187
34 erGlyArgAspArgSerCysAlaLeuProLeuProGlyThrMetGly 50
188 GACGAGTGATCCGAACGACAGAACTCCCTACCGCCAGAGTACATGGGCG 237
51 PheProPhePheGlyGluThrLeuGlnMetValLeuGlnArgArgLysPh 67
238 TTGCGGTTTCAATGGAGAAACGCTCCAGCTGATCTCCAGAGAGGAAGT 287
67 eleuGlnMetLysArgArgLysTyrrGlyPheIleTyrrLysThrHisLeuP 84
288 TCTGGCATGAAACGCGAGAAATACGGGTGCATCTTACAGACGCACTCT 337
84 heGlyArgProThrValArgValMetGlyAlaAspAsnValArgArgIle 100
338 TCGGGAACCCGACCTGTCAAGGGTATGGAGCTGATTAATGTAGCGAGATT 387
101 LeuLeuGlyAspAspArgLeuValSerValHisThrProAlaSerValAr 117
388 CTGCTGGCGGAACACACAGCTGTGTGTCTCAGTGGCCAGCATCTGAG 437
117 gThrIleLeuGlySerGlyCysLeuSerAsnLeuHisAspSerHisL 134
438 AACCATCTGGGCTGTGACACCTCTCCATGTCCATGGAGTTCAACACA 487
134 ySglnArgLysValIleMetArgAlaPheSerArgGluAlaLeuGlu 150
488 AAAACAAAGAAAAAGCCCATTTAGAGGCGTTCTCTCGAGATGCTGTGAG 537
151 CysTyrrValProValIleThrGlnGluValGlySerSerLeuGluInTr 167
538 CACTACATTCCTCGTGTATCCACGAGGAGGTGAAGAGCGCATACAGAAATG 587
167 pLeu.....SerCysGlyGluArgGlyLeuLeuValTyrrProGluV 181
588 GCTGCAAAAAGACTCTCTC.....GTGCTGTTATTCAGAAA 625

```

```

181 aLLysArgLeuMetPheArgIleAlaMetArgIleLeuGlyCysGlu 197
    ::::::::::::::::::::::::::::::::::::::::::::::::::::
626 TGAAGAACTCATGTTTCGATAGCTATGAGAAATCCTCTGGTTTGAA 675
198 Pro...GlnLeuAlaGlyAspGlyAspSerGlnGlnGlnLeuValGluAl 213
    |||
676 CCAAGCAAAATTAAGACGAC.....GAGCAAGAACTGCTGGTGAAGC 716
213 abheGlnGlnMetThrArgAsnLeuPheSerLeuProIleAspValProp 230
    |||
717 TTTTGAGAAATGATCAAAACTTCTCTCTTCGCAATGACGCTCTCT 766
230 heSerGlyLeuThrArgGlyMetLysAlaArgAsnLeuIleHisAlaArg 246
    |||
767 TCAGTGTGTGTACAGGGGTGTGAGGCAACGCAATTCATCTCCCAAA 816
247 IleGlnGlnAsnIleArgAlaLysIleCysGlyLeuArgAlaSerGluAl 263
    |||
817 ATTGAGCAAAATCATCAGACAGCAAAATTCAGAT...GACGCAATGAAAA 863
263 aGlyGlnGlyCysLysAspAlaLeuGlnLeuIleGlnHisSerTrpG 280
    |||
864 CGACAGAAATACAAAGAGCCCTTCAGCTGTGTGATGAGACAGACAGAA 913
280 IaArgGlyLysArgLeuAspMetGlnAlaLeuLysGlnSerSerThrGlu 296
    ::
914 GAAGTACAGCACTTTAGTTGTCGAGGATGAAGAAAGACAGTACAGAG 963
297 LeuLeuPheGlyGlyHisGlnThrThrAlaSerAlaAlaThrSerLeuI 313
    |||
964 CTCTATTGAGGTATGAAACCGCCGACGCTGCACACCTCATCTGT 1013
313 eThrTrpLeuGlyLeuTrpProHisValLeuGlnLysValaArgGlnLut 330
    ::
1014 CAGTTTCTGGGTGACACAGACAGAGGTGTGACAGAGGTACAGAGAGAG 1063
330 euLysSerLysGlyLeuLeuCysLysSerAsnGlnAspAsnLysLeuAsp 346
    |||
1064 TTCAGAGAGAGGTGAAATGGGCATGATACCTCGAAAGGCTGTAGT 1113
347 MetGlnIleLeuGlnLeuLysTrpIleGlyCysValIleLysGlnLut 363
    |||
1114 ATGAGACTGTGTGACAGCTGAAATGACATGATGTGTGATTAAGACAC 1163
363 rLeuArgLeuAsnProProValProGlyGlyPheArgValaLeuLysT 380
    |||
1164 TCCTTAAGATCAACCTCTCTGTCGAGAGATTCAGAGTCCGACTCAAAA 1213
380 hrPheGlnLeuAsnGlyTrpGlnIleProLysGlyTrpAsnValIleLys 396
    |||
1214 CTTTGAATGATGATGTTACCAAAATTCCTAAAGATGAGACGTCATTAC 1263
397 SerIleCysAspThrHisAspValaIleGlnIlePheThrAsnLysGlu 413
    |||
1264 AGCATCTGTGACACGCGATGTGCGCAGCTCTTCCAAACAAAGAGGA 1313
413 upHeAsnProAspArgPheSerAlaProHisProGlnAspAlaSerArgP 430
    |||
1314 GTTCCAGCCGAGAGATTCATGAGCAAAAGCTGTGAGGAGCGGTCCAGT 1363
430 heSerHeiLeuProPheGlyGlyLysArgSerCysValaGlyLysGlu 446
    |||
1364 TTAACTACATCCCTTCGAGAGAGATTCAGGATGTGTGTGGCAAAAG 1413
447 PheAlaLysIleLeuLeuLysIlePheThrValaGlnLeuAlaArgHisC 463
    |||
1414 TTGCGCAAGGTGTACTCAAGATCTTTAGTTAGTTAAGCAGCAGCATGT 1463
463 sAspTrpGlnLeuLeuAsnGlyProProThrMetLysThrSerProThrV 480
    |||
1464 CAATGTGATTTCTCAAAAGGAGCCCGCAATGAAGAGCCCGAGCTA 1513

```

```

480 aLTrProValAspAsnLeuProAlaArgPheThrHisPhe 493
    ::::::::::::::::::::::::::::::::::::::::::::::::::::
1514 TTTACCAAGTGTGACATCTCCCTACCAATTCACCTACTTAT 1554
seq_name: /cgn2_6/ptodata/2/lna/58_COMB.seq:US-08-882-164D-38

```

seq_documentation_block:

```

; Sequence 38, Application US/08882164D
; Patent No. 6306624
; GENERAL INFORMATION:
; APPLICANT: Petkovich, P. Martin, White, Jay A.,
; TITLE OF INVENTION: Beckett, Barbara R., Jones, Glenville
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Blake, Cassels & Graydon
; STREET: Box 25, Commerce Court West
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5L 1A9
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 Mb storage
; COMPUTER: COMPAQ, IBM PC compatible
; OPERATING SYSTEM: MS-DOS 5.1
; SOFTWARE: WORD PERFECT
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/882,164D
; FILING DATE: June 25, 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/667,546
; FILING DATE: June 21, 1996
; APPLICATION NUMBER: 08/724,466
; FILING DATE: October 1, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Hunt, John C.
; REGISTRATION NUMBER: 36,424
; REFERENCE/DOCKET NUMBER: 50767/00010
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 863-4344
; TELEFAX: (416) 863-2653
; INFORMATION FOR SEQ ID NO: 38:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4164 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-882-164D-38

```

alignment_scores:

```

Quality: 1712.00 Length: 1046
Ratio: 4.019 Gaps: 10
Percent Similarity: 40.727 Percent Identity: 38.815

```

alignment_block:

```

US-09-668-482-4 x US-08-882-164D-38 ..
Align seg 1/1 to: US-08-882-164D-38 from: 1 to: 4164

```

```

40 CysAlaLeuProLeuProProGlyThrMetGlyPheProPheGlyGlu 56
    |||
73 TCGGCTCTCTCTCCACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 99
56 uThrLeuGlnMetValLeuGlnArgArgLysPheLeuGlnMetLysArg 73
    |||
100 .....CAGCGAGAGAAATGTTCTGCAGATGAAGCGCA 130
73 rGlyTrpGlyPheIleTrpGlyLysThrHisLeuPheGlyArgProThrVal 89
    |||
131 GGAATACGAGCTTCACTACAGACGACATCTGTGTGGCGGCGCAGGTG 180
90 ArgValMetGlyAlaAspAsnValaArgArgIleLeuLeuGlyAspAsp 106

```

106 gLeuValSerValHisTrpProAlaSerValArgThrIleLeuGlySerG 123
231 GTTGGTGTGGGTCAGTGGCCGGTGGTGGTGGTGGTGGTGGTGGTGG 280
123 lYcysLeuSerAsnLeuHisAspSerHisIstysGlnArgLys..Lys.. 138
281 GGTGCTCTCCACCTGCAGCATTCCTCTCGCACAAGCAGCAGCAAGAG 330
138 138
331 AGGCTGAGCTGGCACTCTTGGCTGGCAGGAGAGACCTCATCTATGCT 380
138 138
381 TGGTTCAGGCAAAATAGAAATGCGGGGCGAGGGCTAGTCTATGTGTGG 430
138 138
431 GACGAGACCTCTCTCTATCTGATCCACTTTAGCTTTCTGCTAGCAGC 480
138 138
481 TGGGTTAGTCCCTGGGGGGGACTGAAATTCTTGAAGGGTACTCGGAAAG 530
138 138
531 CGAAGGGGGGGGGCTGAGGAAAGTAGAGATTGTAAACCTCTCTCTC 580
139 139
581 CTGGGGGGTGGCTCAGTGTATTATGAGGCTTTCAGCCGCGAGCAGCTCA 630
150 uCytyrValProValIleThrGluGluValGlySerSerLeuGluGln 167
631 GTGCTACGTGCGCGTGTGATGCTGAGGAGTCAAGAGTTGTCTGGAGCAG 680
167 rPLeuSerCysGlyGluArgGlyLeuLeuValTyrProGluValLysArg 183
681 GGCTAAGCTGGCGGCGAGCGCGCTCTGCTGCTACCCCGAGGTGAACGC 730
184 LeuMetPheArgIleAlaMetArgIleLeuLeuGlyCysGluProGln 200
731 CTCATGTTCCGATGCCATGCGATCTGCTGGCTGCGAGCGGCTCC 780
200 uAlaGlyAspGlyAspSerGluGlnGlnLeuValGluAlaPheGluGln 217
781 AGCGGGCGCGGCGAGGAGCAGCAGCAGCTCGTGAAGGCTTTCAGAGGA 830
217 eTThrArgAsnLeuPheSerLeuProIleAspValProPheSerGlyLeu 233
831 TGACCCGCAATCTCTCTCTCTCCATGACGTGCCCTTTAGGGGCTG 880
234 TyrArg..... 235
881 TACCGGGTAAGGGCGTTTGGCGAGTCGAGTAGGGAACGCAAGCTCG 930
236 Gly..MetLysAlaArg 240
931 GCATCCGCTCAACGCCACGCTCTCTCCGCGCTCAGGGGTGAAGGCGCG 980
241 AsnLeuIleHisAlaArgIleGluGlnAsnIleArgAlaLysIleCysG 257
981 AACCTTATACAGCGCGCATCGAGAGAACTTCGCGCAAGATCCGCGC 1030
257 yLeuArgAlaSerGluAlaGlyGlnGlyCysLysAspAlaLeuGlnLeu 274
1031 GCTTACGCTTACAGCGCGATGGGGGTTCAGAGGAGCGCGCTGAGCTCC 1080
274 euIleGluHisSerTrpGluArgGlyGluArgLeuAspMetGln..... 288

1081 TGATTGACACTCTGTGGAGAGGGAGAGAGAGCTGATATGACGTGAGA 1130
288 288
1131 AGCAATTTCAAAGGTGCCAAGGCGGGGAGTGCCTCTACTTTCCAGA 1180
288 288
1181 CACACTTCTGGGGTCTCCAAAGCCGTGCAAGGCCACAGTACTTCCA 1230
288 288
1231 GTGGCGCGGATGCTAGGTCTAGAGCTTTTCAACCTGTGGTGTGACCC 1280
288 288
1281 CTTCAGGAGCCAAACACCTTTCAGAAAGGTGCGCTAAGAGCATCTGC 1330
288 288
1331 ATATCCGATATTACATCAGAAACATACATGACAAATTCGCTTAT 1380
288 288
1381 GAAGTAGCAACAAGATATTTTATCGTTGGGGGTCCACACACAGAGG 1430
288 288
1431 AACCGTATTAAAGGTGGCATTTGCTAGAGAGCTGTGAGAGGGGTGGC 1480
288 288
1481 TGAGCAATGGGGAAGATCCCAAGTTCAAAGGCAAGGCTCATCTACAA 1530
288 288
1531 GGTAAAGGAGAGAGAGCATTAAGGAGTTTGGCTTTTGTGTGCG 1580
288 288
1581 TCTGTATTAACCTTCGTTTATCCCTTACCTTGAAGCTTCGTCG 1630
288 288
1631 CAATTTCTCAGTCAGGTCTTTTACAGTGCATATAATGACACACA 1680
289 289
1681 TCTGTATTAACCTTCGTTTATCCCTTACCTTGAAGCTTCGTCG 1730
296 GluLeuLeuPheGlyGlyHisGluThrThrAlaSerAlaIleThrSer 312
1731 GAGCTCCCTTTTGGTGTGATGAACTACAGCCAGCTGCTCGACGTCAC 1780
312 uIleThrTyrLeuGlyLeuTyrProHisValLeuGlnLysValArgGlu 329
1781 GATCATTACCTAGGACTCTACCCACATGTCTCCAAAGGTTCCGAAG 1830
329 LuLeuLysSer..... 332
1831 AGATTAAGCAAGTAGATGATTTAGAGTTCCCATTTGCTTACAGA 1880
332 332
1881 CATTCCTATTAAACACACACACACCCCACTGTATATAGTTTGTCTC 1930
332 332
1931 GATACACCACTACTAGACAGTGAAGATCTGAGAGTAGTGGAGCTGT 1980
332 332
1981 GGGGAGAGACTCCACCTCTGTGAATTTAAAGGCACTGTGTGTACTGG 2030

Align seg 1/1 to: US-08-882-164D-36 from: 1 to: 2677

```
1 MetGlyLeuProAlaLeuLeuAlaSerAlaLeuCysThrPheValLeuR 17
|||||
676 ATGGGGCTCCCGCGCTGCTGGCCAGTGGCTGCTGCACCTTGCTGCTGCC 725
17 OLeuLeuLeuPheLeuAlaAlaIleLysLeuTrpAspLeuTyrCysValS 34
|||||
726 GCTGTGCTCTCTCTCGGCTGCATCAAGCTTGGGACCTGTACTGGCTGA 775
34 erGlyArgAspArgSerCysAlaLeuProLeuProGlyThrMetGly 50
|||||
776 GCGGCGCGGACCGCAGATTGTGCCCTCCCATTTGCCCGCGGACTATSGGS 825
51 PheProPhePheGlyGluThrLeuGlnMetValLeu.Gln..... 63
|||||
826 TTCCCTCTTTGGGGAACCTTGCAGATGNTACTNCAAGTAAAGGAGGG 875
63 ..... 63
876 TGGGGCGGGACAGGCTGCTTCCCGGAGCGCGGCTGCTGGGCTTCT 925
63 ..... 63
926 GCTGAAGTCGGGGTAGGGCCCCCGGAGAGCATGCTATTGCGGTAGAG 975
63 ..... 63
976 CAGGGCTGGCGGAGCGCGGCTCCCGGMYKSCTCAMGSCRCRWK 1025
64 ..... ArgArgLysPheLeuGlnMetLysArg 72
1026 TMCCTCCGCCCTYMCCTCCCAMAGCGGARSAAARWKYKGRGATGAAGCC 1075
73 ArgLysTyrGlyPheIleTyrLysThrHisLeuPheGlyArgProThrVa 89
|||||
1076 AGGAATTCGGCTTCATCTACAGACGATCTGTGCGGCGGCCACCGT 1125
89 LArgValMetGlyAlaAspAsnValArgArgTlleLeuLeuGlyAspAspA 106
|||||
1126 ACGGGTGTGGGCGCGGACATGTGCGGCTCTTGGCTCGGAGAGACCC 1175
106 rGLeuValSerValHisTrpProAlaSerValArgThrIleLeuGlySer 122
|||||
1176 GCGTGTGTCGCTCCATCGCCAGCGCTGGTGGCCACCATTCCTGGGACT 1225
123 GlyCysLeuSerAsnLeuHisAspSerSerHisLysGlnArgLysLys.. 138
|||||
1226 GGTGTGCTCTTAACCTGACGACTCTGACACAGCAGCGCAAGAGGT 1275
138 ..... 138
1276 GGGGCGAGAGAGCGACGCGCTGGACAGGAGGGGACCCCATTTATGAGCG 1325
138 ..... 138
1326 GAATTCGGCGATGATGATGATGAGCGCGGCTAGCAGCTTGAAGTGGGCT 1375
138 ..... 138
1376 AGGACCTCTGCCAGCTCCAGTTAGCTTTCCAGCTCGGAGAGTGCAT 1425
138 ..... 138
1426 GTGTGTGACAGAGACTGGGGGTGTCTGGAAGGGAGCGGCGTAAAGCAGAG 1475
139 ..... ValIle 140
1476 GGGCGGATGAGAGCTTTTAACGCTGCTCCCTCTCGGAGACTCAGGTGATT 1525
141 MetArgAlaPheSerArgGluAlaLeuGluCysTyrValProValIleTh 157
|||||
```

```
1526 ATGGCGGCTTCAAGCCGAGAGACTGCAATGTACTGCGGTATTCAC 1575
157 rGluGluValIleGlySerSerLeuGlnTrpLeuSerCysGlyGluArg 174
|||||
1576 CGAGGAAGTGGGCGAGCAGCCTGAGACAGTGGCTGAGACTCGCGGACGGG 1625
174 lYLeuLeuValTyrProGluValAlaLysArgLeuMetPheArgIleAlaMet 190
|||||
1626 GCCTCTGTGTTACCCCGAGGTGAAGCGGCTCATGTTCCGAATCCGCATG 1675
191 ArgIleLeuLeuGlyCysGluProGlnLeuAlaGlyAspGlyAspSerG 207
|||||
1676 CCATCTCTACTGGGCTGCGAACCCTGAGCGGGGAGCGGAGACTCCGA 1725
207 uGlnGlnLeuValGluAlaPheGluGluMetThrArgAsnLeuPheSerL 224
|||||
1726 GCAGCAGCTTGTGAGGCGCTTCGAGGAATGACCCGCAATCTCTCTCCG 1775
224 euProIleAspValProPheSerGlyLeuTyrArg..... 235
|||||
1776 TGCCTATGACAGTGCCTTCAAGGGGCTTACCAGGTAAGGCGGCAAC 1825
235 ..... 235
1826 GGGCTGCGGACTAGGGGCGGGACCTGGGCGTGTGCTCACCGCGCGCG 1875
236 ..... Gly.MetLysAlaArgAsnLeuIleHisAlaArgIle 247
1876 CTCTCTGCGCTCAGGGGCATGAAGGCGGGAACTCATTCACGCGGCATC 1925
248 GluGlnAsnIleArgAlaLysIleCysGlyLeuArgAlaSerGluAlaG 264
1926 GAGCAGAAATTCGCGCCAGATCTGCGGCTGCGGCTCCGAGAGCGGG 1975
264 yGlnGlyCysLysAspAlaLeuGlnLeuLeuIleGluHisSerTyrGluA 281
|||||
1976 CCAGGGCTGCAAAAGACGCGCTGACAGCTGTGATGAGCAGCTGCTGGAGA 2025
281 rGlyGluArgLeuAspMetGln..... 288
2026 GGGGAGAGGGCTGGACATGACGTGAGTAGCAGCTTCAGACCAGGCACT 2075
288 ..... 288
2076 GCGGAGTTTGTGTCCTGCGCTTCCAAAGCGCTGTTCCTGGGGCCCCCA 2125
288 ..... 288
2126 AGCGGCGCTGGGGCCAGACTTCTGAGTGGGCGCGGCTCAGACTA 2175
288 ..... 288
2176 CAGCTATGGAATCCCGAAGGAAGCTGAGACACCGGTCAGAGAGCTGC 2225
288 ..... 288
2226 GGAAGGGCTGCGGGAAGCTGGAGCATCCCTAGCCTTAMCAGGTT 2275
288 ..... 288
2276 TCAAAGGAAGTTGGAATTTGCAAAAATGTTAATAAGAACCTTGCAT 2325
288 ..... 288
2326 TTTAATAAAGCTTAAGACTTTAAGCAGAGTTTCGGTAAAGCGGGGTGC 2375
288 ..... 288
2376 TACTGCGCTTACTGCTCCAGCTGAAGCTAAAGGAGCTTGCAATTTGTTTA 2425
288 ..... 288
2426 AAGATATTGCTTCTCTTGTGACTTCTGTGACGAAAAACATTTAGCCCTTCTA 2475
```

288 288
2476 GTCTTCCCTCCAGAACTCTCATGTTGATCTGAGTAATCCTTGTGCAA 2525
289 Ala..Leu 290
2526 CCCGAGCAGACTTGTGAGATGGGTCTACACTTATTCTTAGGCACTA 2575
291 LysGlnSerSerThrGluLeuLeuPheGlyGlyHisGluThrThrAlaLeu 307
2576 AAGCAATCTTCAACCGAATCCTCTTGTGAGGAGACAGAAACGAGGCGAG 2625
307 TAlAlaIhSerIeulhThrIhThrLeuGlyLeuTyrrProHisValLeuG 324
2626 TGCAGCCACATCTCTGATCTACCTACCTGGGGCTCTACCCACATGTTCTCC 2675
324 In 324
2676 AG 2677

seq_name: /cgn2_6/ptodata/2/lna/6A_COMB.seq:US-08-724-466B-11

seq_documentation_block:
; Sequence 11, Application US/08724466B
; Patent No. 6063606
; GENERAL INFORMATION:
; APPLICANT: Petkovich, P. Martin, White, Jay A.,
; APPLICANT: Beckett, Barbara R., Jones, Glenville
; TITLE OF INVENTION: Retinoid Metabolizing Protein
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Blake, Cassels & Graydon
; STREET: Box 25, Commerce Court West
; CITY: Toronto
; STATE: Ontario
; ZIP: M5L 1A9
; COUNTRY: Canada
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 Mb storage
; OPERATING SYSTEM: MS-DOS 5.1
; SOFTWARE: WORD PERFECT
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/724,466B
; FILING DATE: October 1, 1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/667,546
; FILING DATE: June 21, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Hunt, John C.
; REGISTRATION NUMBER: 36,424
; REFERENCE/DOCKET NUMBER: 50767/00004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 863-4344
; TELEFAX: (416) 863-2653
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 351 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-724-466B-11

alignment_scores:
Quality: 614.00 Length: 117
Ratio: 5.248 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-668-482-4 x US-08-724-466B-11 ..
Align seg 1/1 to: US-08-724-466B-11 from: 1 to: 351

296 GluLeuLeuPheGlyGlyHisGluThrThrAlaSerAlaAlaThrSerLe 312
1 GAACTCTCTTTGGAGACACGAAACACAGGCGCAGTGCACCATCTCT 50
312 uIleThrTyLeuGlyLeuTyrrProHisValLeuGlnTyValArgGluG 329
51 GATCATTACTGCTGGGCTCTACCCACATCTTCTCCAGAAAGTGCAGAG 100
329 LuLeuLysSerLysGlyLeuLeuGlySerAsnGlnAspAsnLysLeu 345
101 AGCTGAAGATGAGGGTTTACTTTCGCAAGACATCAACACAAAGTTG 150
346 AspMetGluLeuLeuGlnLeuLysTyrrIleGlyCysValIleLysG 362
151 GACATGAATTTTGGACACACTTAATATCATCGGTGTATTATAGA 200
362 uThrLeuArgLeuAsnProProValProGlyGlyPheArgValAlaLeu 379
201 GACCTTGCACATGCCCATCTCCAGAGGTTTCGGTGTCTGA 250
379 yThrPheGluLeuAsnGlyTyrrGlnIleProLysGlyTrpAsnValIle 395
251 AGACTTTGATTAATGATGATACCAATTCCCAAGGCTGGAATGTTATC 300
396 TyrSerIleCysAspThrHisAspValAlaGluIlePheThrAsnLysG 412
301 TACAGTACTGTGATCTCATGTGATGGCAGAGATCTTACCCACAGAGA 350
412 u 412
351 A 351

seq_name: /cgn2_6/ptodata/2/lna/6B_COMB.seq:US-08-882-164D-11

seq_documentation_block:
; Sequence 11, Application US/08882164D
; Patent No. 6306624
; GENERAL INFORMATION:
; APPLICANT: Petkovich, P. Martin, White, Jay A.,
; APPLICANT: Beckett, Barbara R., Jones, Glenville
; TITLE OF INVENTION: Retinoid Metabolizing Protein
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Blake, Cassels & Graydon
; STREET: Box 25, Commerce Court West
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5L 1A9
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 Mb storage
; OPERATING SYSTEM: MS-DOS 5.1
; SOFTWARE: WORD PERFECT
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/882,164D
; FILING DATE: June 25, 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/667,546
; FILING DATE: June 21, 1996
; APPLICATION NUMBER: 08/724,466
; FILING DATE: October 1, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Hunt, John C.
; REGISTRATION NUMBER: 36,424
; REFERENCE/DOCKET NUMBER: 50767/00010
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 863-4344
; TELEFAX: (416) 863-2653
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 351 base pairs

```

;
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-882-164D-11

```

```

alignment_scores:
  quality: 614.00      length: 117
  ratio: 5.248         gaps: 0
  percent similarity: 100.000  percent identity: 100.000

```

```

alignment_block:
US-09-668-482-4 x US-08-882-164D-11 ..

```

```

Align seg 1/1 to: US-08-882-164D-11 from: 1 to: 351

```

```

296 GULeuLeuPheGlyGlyHisGluThrThrAlaSerAlaAlaThrSerle 312
1 GACACTCCTCTTGGAGACACACAAACCGCCAGTGCAGCCACATCTCT 50
312 uileThrTyrlleuGlyLeuTyrlProHisValleuGlnValArgGluG 329
51 GATCACTTACTGGGCTCTACCCACATGTCTCCAGAAAGTGGCAGAG 100
329 LuLeuLysSerIysGlyLeuLeuGlyLysSerAsnGlnAspAsnLysLeu 345
101 ACCTAAGAGTAAAGGCTTACTTTGCAAGAGCAATCAAGACAACAGTGG 150
346 AspMetGluIleLeuGlnLeuGlnLeuLysTyrlleGlyCysValIleLysG 362
151 GACATGGAATTTTGGACACACTTAATATCATCGGTGTATTATAGGA 200
362 uThrLeuArgLeuAsnProProValProGlyGlyPheArgValAlaLeuL 379
201 GACCTTCGACTGAATCCCGCATTCACAGAGGGTTCCGGTTGCTCTGA 250
379 ySThPheGluLeuAsnGlyTyrlGlnIleProLysGlyTyrlProAsnValIle 395
251 ACACCTTGAATTAATGATACCAAGATTCCCAAGGGCTGGAATGTATTC 300
396 TyrSerIleCysAspThrHisAspValAlaGluIlePheThrAsnLysG 412
301 TACAGTATCTGTGATACATCATGATGTGGCAGAGATCTTACCAACAAGA 350
412 u 412
351 A 351
seq_name: /cgr2_6/ptodata/2/ina/5B_COMB.seq:US-08-622-166A-1
seq_documentation_block:
; Sequence 1, Application US/08622166A
; Patent No. 5952545
; GENERAL INFORMATION:
; APPLICANT: KONCZ, CSABA
; APPLICANT: MATHUR, JAIDEEP
; APPLICANT: SZEKERES, MIKLOS
; APPLICANT: ALTMANN, THOMAS
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING CYTOCHROME
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch, LLP
; STREET: P.O. Box 747
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

```

```

;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/622,166A
; FILING DATE: 27-MAR-1996
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: SVENSSON, LEONARD R.
; REGISTRATION NUMBER: 30,330
; REFERENCE/DOCKET NUMBER: 0147-0153P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; TELEX: 248345
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1608 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: A. thaliana
; IMMEDIATE SOURCE:
; LIBRARY: lambda gt10
; CLONE: C204
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 48..1466
US-08-622-166A-1

```

```

alignment_scores:
  quality: 433.50      length: 486
  ratio: 1.521         gaps: 13
  percent similarity: 58.642  percent identity: 26.543

```

```

alignment_block:
US-09-668-482-4 x US-08-622-166A-1 ..

```

```

Align seg 1/1 to: US-08-622-166A-1 from: 1 to: 1608

```

```

19 LeuLeuPheLeuAlaAlaIleLysLeuTrpAspLeuTyrlCysValSerle 35
66 CTCCTCTCTCTCTTCATCGCGCCGCTTCCTCTCTCTCTCTCTCCG... 113
35 yArgAspArgSerCysAlaLeuProLeuProProGlyThrMetGlyPhe 52
114 .CGTACAGCTTACCGTCGAGATGGCTGTGCTCGCGGAGACCTTGCTCG 162
52 roPhePheGlyGluThrLeuGlnMetVal.....LeuGln 63
163 CTCCTGATAGAGAGACTTTTCAGCTGATCGAGCTTACAAACAGAGAAC 212
64 ATGATGlyPheLeuGlnMetLysArgArgLysTyrlGlyPheIleTyrl 80
213 CCGAGCCCTTCAATCGACGAGAGAGTACCGCGGTACGGTTGCTTCAT 262
80 sThrHisLeuPheGlyArgProThrValArgValMetGlyAlaAspAsnV 97
263 GAGCATCTTTTGGGAACCGACGATTTTTCAGCTGACCGCGAAGACA 312
97 aLArgArgIleLeuLeuGlyAspAspArgLeuValSerValHisTrpPro 113
313 ACCGGTTGTCTCTCAGAACGAAAGCACTTTTGAAGTTCCTTATCTCT 362
114 AlaSerValArgThrIleLeuGlySerGlyCysLeuSerAsnLeuHis 130
363 GCTTCATTGTACCTTTTGGGAGAACACTCTCTGCTTCTTATGAAAG 412
130 pSerSerHisLysGlnArgLysLysValIleMetArgAlaPheSerArg 147
413 TTCTTTGATTAACGATACACTCTCTACCATGAGCTTGTGAATCTT 462

```

```

147 LuAlaLeuGluCysTyrValProVal...IleThrGluGluValGlySer 162
    :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
463 CAATCATTTAAGACCATTCATGCTTGAATATTGACCGGTTAGTCCGGTTT 512
163 SerLeuGluGlnThrLeuSerCysGlyLysGlyLeuLeuValTyrPr 179
    :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
513 AACTTGATCTTGCTCTCT...CGTGTCTCTCTCATGSA 550
179 ogLuValLysArgLeuMetPheArgIleAlaMetArgIleLeuLeuGlyC 196
    :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
551 AGAAGCCAAAAGATAGCTTGTGACCTAACGGTGAAGCAGTTGATGAGCT 600
196 ysgLuProGlnLeuAlaGlyAspGlyAspSerGluGlnGlnLeuValGlu 212
    :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
601 TTGATCCA.....GGGGAATGGAGTGAGATTAAAGGAA 635
213 AlaPheGluGluMetThrArgAsnLeuPheSerLeuProIleAspValPr 229
    :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
636 GAGTATCTTTTGTCTCATCGAAGCTTCTCTCTCTCTCTCTCTCTCTC... 683
229 oPheSerGlyLeuTyrArg...GlyMetLysAlaArgAsnLeuIleHisA 245
    :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
684 .TTCTCACACATTACCGCAAAACCATCCACAGCGCGAGG..... 722
245 laArgIleGluGlnAsnIleArgAlaLysIleCysGlyLeuArg..... 259
723 ..AAGGTGGCGAGGCGGTGACGCGTGTGTGATGAAAGGAGGAGGAG 770
260 AlaSerGluAlaGlyGlnGlyCysLysAspAlaLeuGlnLeuLeuIle.. 275
771 GAGAGAAAGAGAGCGAGAGAAAGAAAGATATCTTCCGCGCTTGCTTC 820
276 ....GluHisSerTrpGluArgGlyLysArgLeuAspMetGlnAlaLeu 291
821 GCGGAGATGATGATTTCCGATGAGAGATGTTGACTTC..... 860
291 ysgLnsSerSerThrGluLeuLeuPheGlyGlyHisGluTrpHisAlaSer 307
861 .....TTGGTGGCTTACTTGTCCGCGTTATGAAACCACTCCACG 902
308 AlaAlaThrSerLeuIleThrTyrLeuGlyLeuTyrProHisValLeuG 324
903 ATCATGACTCTCGCCGCAATTTCTACCGAGACTCTTACTCTCTGTC 952
324 nLysValArgGluGlnLeuLysSerLysGlyLeuLeuCysLysSeransg 341
953 TCAACTCAAGGAAGAG.....CATGAAAGATTAGGCAATGAGAGAGTG 996
341 LnsAspAsnLysLeuAspMetGluIleLeuGlnGlnLeuLysTyrIleGly 357
997 ATTGCTTAGCTTGAATGGAGTGATTACAAAGTCAATGCCATTCACAA 1046
358 CysValIleLysGluThrLeuArgLeuAsnProProValProGlyGlyPh 374
1047 TGTGTGTTAATGAGACGCTACGAGTGGCTAACATCATCGCGGTGTTT 1096
374 earGValAlaLeuLysThrPheGluLeuAsnGlyTyrGlnIleProLysG 391
1097 CAGACGCGCAATGACGAGTGTGAGATCAAAAGCTTATTAATTCAAAG 1146
391 LysTrpAsnValIleTyrSerIleCysAspThrHisAspValAlaIle 407
1147 GGTGGAAGTATTTCTCATCGTTTACACCGGCTCATTTAGACCCCAAC 1196
408 PheThrAsnLysGluGluPheAsnProAspArgPheSerAlaProHisPr 424
1197 TTCAAGATGCTGCGACTTCAACCTTGGAGATGGCAGACCACTCGGT 1246
424 ogLuAspAlaSerArgPheSerPheIleProPheGlyGlyLysLeuArgS 441
1247 AACGACAGGCCCTTCTAATGTTCACACCGTTTGGAGGCGCAAGGC 1296

```

```

441 erCysValGlyLysGluPheAlaLysIleLeuLysIlePheThrVal 457
    :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
1297 TATGTCGCCGTTACGAGCTGCGTACGGTTGACACTCTCTGTTTCTTCAC 1346
458 GluLeuAlaArgHisCysAspTrpGlnLeuLeuAsnGlyProProHisMe 474
    :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
1347 CGCCTAGTACAGCGCTCAGTTGG.....GTTCCGACAGA 1381
474 LysThrSerProThrValTyrProValAspAsnLeuProAlaArgPheT 491
1382 GCMAACACAGCTGTTTCTTCCAACTACAAAGACGCAAGAACGGTACC 1431
491 hrHisPhe 493
1432 CGATCTTC 1439

```

```
seq_name: /cgn2_6/prodata/2/lna/66_COMB.seq:US-08-991-677-3
```

```

seq_documentation_block:
: Sequence 3, Application US/08991677A
: Patent No. 6252135
: GENERAL INFORMATION:
: APPLICANT: Chianq, Vincent L
: APPLICANT: Carraway, Daniel T
: APPLICANT: Smeltzer, Richard H
: TITLE OF INVENTION: Production of Syringyl Lignin in Gymnosperms
: FILE REFERENCE: 50617
: CURRENT APPLICATION NUMBER: US/08/991,677A
: CURRENT FILING DATE: 1997-12-16
: EARLIER APPLICATION NUMBER: US 60/033,381
: EARLIER FILING DATE: 1996-12-16
: NUMBER OF SEQ ID NOS: 11
: SOFTWARE: patentIn Ver. 2.0
: SEQ ID NO 3
: LENGTH: 1883
: TYPE: DNA
: ORGANISM: Liquidambar styraciflua
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (74)..(1606)
US-08-991-677-3

```

```

alignment_scores:
Quality: 348.50 Length: 507
Ratio: 1.267 Gaps: 22
Percent Similarity: 34.241 Percent Identity: 26.824

```

```
alignment_block:
US-09-668-482-4 x US-08-991-677-3 ..
```

```
Align seg 1/1 to: US-08-991-677-3 from: 1 to: 1883
```

```

14 PheValLeuProLeuLeuLeuPheLeuAlaIleLysLeuTrpAspLe 30
    :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
125 TTATATATACCTTGTCTACTCTATATGGCGCTAGTATCTCGGCTT... 169
30 uTyrCysValSerGlyArgAspArgSerCysAlaLeuProLeuProG 47
170 .....CGCCAGAGA.....CTACCATACCCACGAG 194
47 LysThrMetGlyPheProPhePheGlyGluThrLeuGlnMetValLeuGln 63
195 GCCCAAAAGGCTTACCGGTATGAGGAACATGCTCATGATGATCAATC 244
64 ArgArgLysPheLeuGlnMetLysArgArgLysTyrGlyPheIleTyrI 80
245 ACTACACGAGAGCTCGCAAACTCGCCAAACAATAGCGGCTTATTTCA 294
80 sTrnHisLeuPheGlyArgProThrValArgValMetGlyAlaAspAsnV 97
295 CCTCAAGATGGATTCTTACACATGTTGGCGCTTCCACACCCGACATGG 344

```

```

97  aATgArgIleuLeuGluYAspAspArgLeuValSerValHisTrpPro 113
   :::::||||| :::::||||| :::::|||||
345  CTGGCAAGTCTCTCAAGTCCAGACAAACATCTTCTCG...AACGGCCA 391
114  ALaSerValArgThrIleLeuGlySerGlyCysLeuSerAsnLeu...Hi 129
   |||:::||||| :::::|||||
392  GCCACCAATA.....GCCATCAGCTACCTACCTA 420
129  sAspSerSer.....HisLysG 135
   ::||| :::::
421  TGACCGAGCCGACATGGCTTCCTCCTACGCGCCGCTTTGGCGTCMA 470
135  InArgLysLysValIleMetArgAlaPheSerArgLysAlaLeuGlyCys 151
   ::|||::: ::|||::: ::|||:::
471  TGGTAAACTCTGCTCATGAATATTATTAGCCGGAAGACGACGATCG 520
152  Tyr.....ValProValIleThrGly 158
   ::::: ||| |||:::
521  TGGGAGTCGTCGAGACGAGTCGCTCGGCAATGAGTGTGCGCTC 570
158  uGluValGlySerSerLeuGluGlnTrpLeuSerCysGlyGluArgGlyL 175
   :::::|||||::: :::::|||||
571  CAATATGGGTGAGAG.....GTGAATTCGGCGAG..... 601
175  euLeuValTyrProGluValLysArgLeuMetPheArgIleAlaMetArg 191
   |||||::: ::|||::: ::|||:::
602  ..CTGGTTTTCCTCTGACGAAGAATATTACTTACAGCGCGCTTTGGG 649
192  IleuLeuGlyCysGluProGlnLeuAlaGlyAspGlyAspSerGlu 208
   ::::: :::::
650  ACGATC.....TCGCATGAGGACCAAGA 672
208  nGluLeuValGluAlaPheGluGluMetThrArg.....AsnLeuPheS 223
   :::::||||| :::::||||| :::::|||||
673  CGAGTTCGTGGCCATATGCAAGAGTTTTCGCGCTGTTGGTCTTTTA 722
223  erLeuProIleAspValProPhe.....SerGlyLeu... 233
   ::::: ::|||::: ::|||:::
723  ATATATGCTGATTTTATCCCTTGCTCAAAATGGGTCCTCAGGGGATTAAC 772
234  TyrArgGlyMetLysAlaArgAsnLeuIleHisAlaArgGlyLeuGlnAs 250
   ||| |||||::: ::::: |||||:::
773  GTCAGGCGCAACAAGCGACGAGGGCGCTGATGGGTTTATGACAAAGT 822
250  nIleArgAlaLysIle.....CysGlyLeuArgAlaSerGluAlaGlyG 265
   ||| ||| ::::: |||||:::
823  CATGACGATCATATACAGAGGGGAGTAAACCTCGAGGAGGTGATA 872
265  InGlyCysLysAspAlaLeuGlnLeuIleGlnHisSerTrpLysArg 281
   ||| ||| :::::
873  CTGATATGCTGATGATTTACTCTCTTTTACGGTGAGGAAAGCCAAAGTA 922
282  GlyGluArgLeuAspMetGln...AlaLeuLys..... 291
   ::||| |||||::: ::|||:::
923  AGCGAATGTACGATCTCAAAATTCATCAAACTCACCAAGACACAT 972
292  GlnSerSerThrGluLeuLeuPheGlyGlyHisGluThrThrAlaSerA 308
   ::::: ::|||::: ::|||:::
973  CAAGGCTATCATGAGCTAATGTTTGGAGGACGCAAGCGGCGCTCG 1022
308  LaaIleThrSerLeuIleThrTyrLeuGlyLeuTyrProHisValLeuGln 324
   ||| ||| ::|||::: |||||:::
1023  CGATTGAATGGCCATGACGAGCTGATGAAGAAAGCCGAGAAATTTAAG 1072
325  LysValArgGluGluLeuLys...SerLysGlyLeuLeuCysLysSerAs 340
   |||||::: ||||| |||||
1073  AAGGTCCACAAGAACTCGCGGTGGTGATCTT..... 1108
340  nGluAspAsnLysLeuAspMetGluIleLeuGluGlnLeuLysTyrIleG 357
   |||||::: ::|||::: ::|||:::
1109  ....GACCGGGAGTCTGAAAGAAAGACTTCGAGAAAGCTCACCTACTGA 1154
357  LysValIleLeuGlnThrLeuArgLeuAsnProProValProGlyGly 373

```

```

1155  AATGGTACTGAAGAAATCTTCGCTCCACCAACCATCCACTCCTC 1204
374  PheArgValAlaLeuLysThrPheGluLeuAsnGlyTyrGlnIleProLy 390
   ::|||::: ::|||::: ::|||:::
1205  CTCACGAGACACCGGAGACCGCGAGTGGCGGCTTACTTACCTCCGGC 1254
390  sGlyTrpAsnValIleTyrSerIleCysAspThrHisAspValAlaGluI 407
   ::|||::: ::|||:::
1255  GAATACCGCGGTGATGATCAACGCTGCCCATCGCGCGGACAAAGACT 1304
407  LePheThrAsnLysGluGluPheAsnProAspArgPheSerAlaProHis 423
   ::|||::: ::|||::: ::|||:::
1305  CSTGGGCGGACCAAGATACGTTAGCCCTCCAGGTTTCTCAAGACGCT 1354
424  ProGluAsp.....AlaSerArgPheSerPheIleProPheGlyGly 438
   ||| |||||::: ::|||:::
1355  GTGCCGATTTCAAGGGAACACTTCGATTCATCCATTCGCGGACG 1404
438  YLeuArgSerCysValGlyLysGluPheAlaLysIleLeuLeuLysIleP 455
   ||| ||||| ||| ::|||:::
1405  TCGTGGCTTTCGCCGGTATGCAACTCGGACTCAACGGCTAGAGAGC 1453
455  heThrValGluLeuAlaArgHisCys...AspTrpGlnLeuLeuAsnGly 470
   ||||| ||||| |||||
1454  ..ACTGTGCTTCACCTCTCTTCACTGTTTCACGCTGGAGTTG..... 1492
471  ProProThrMetLysThrSer 477
   ||| ||||| |||
1493  CCGGACGGGATGAACCGAGT 1513

```

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-103-840A-2

seq_documentation_block:

```

; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103, 840A
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

```

alignment_scores:

```

Quality: 344.00 Length: 446
Ratio: 1.376 Gaps: 12
Percent Similarity: 56.054 Percent Identity: 25.112

```

alignment_block:

US-09-668-482-4 x US-09-103-840A-2/rev ..

Align seg 1/1 to reverse of: US-09-103-840A-2 from: 1 to: 4403765

```

32 CysValSerGlyArgAspArgSerCysAlaLeuProLeuProGlyLys 48
||| ||| |||::: ||| |||
4120837TCGGGTTGCGCGGAGCAACGCGCGGATGCTTTGGCGGCGCCACCGGCT 4120788

```



```

147 CACGAGTTGGCCCATCATAGGACACATGTTAATGATGGACCAACTCACC 196
|||||
65 ArglySPheLeuGlnMetlySArgAlSlyTYrGlyPheIleTyrLys 81
|||||
197 CACCGTGCTTACCAATTAGCTAAAGATAGCGGATTTGGCCATCT 246
|||||
81 HisLeuPheGlyArgProThrValAlrValMetGlyAlaAspAsnVala 98
|||||
247 CCGCATGGGATTCCTCCATATGATACCTGCTCATCACCAGGAGTGCTC 296
|||||
98 rgrArgIleLeuLeuGlyAspAspArgLeuValSerValHisTPrProAla 114
|||||
297 GACAAAGTCTTCAAGTCCAGACAGCGCTCTTCG...AACGGCCCTGCA 343
|||||
115 SerValArgThrIleLeuGlySerGlyCysLeuSerAsnLeu...HisAs 130
|||||
344 ACTATA.....GCTATAAGCTATCTGACTTACGA 372
|||||
130 pSerSer.....HisLysGlnA 136
|||||
373 CCGAGCGGACATGGCTTTCGCTACTACGACCGTTTGGAGACAGATGA 422
|||||
136 rglYslySValIleMetArgAlaPheSerArgGluAlaLeuGlyCysTyr 152
|||||
423 GAAAGTGTGTGATCATAGAGTGTTAAGCCGTAAGAGCTGATCATG 472
|||||
153 ValProVal.....IleThrGluGluValGlySerS 163
|||||
473 GCTTCATGCTGTGATGAAAGTGACAAATGTCGGTGCTCTTGTA 522
|||||
163 rLeuGluGlnTTrpleuSerCysGlyAlaGlyLeuValTyrProG 180
|||||
523 CGTTGTAAACCTATAAAGTCGGGAGCAA.....ATTGTCAC 563
|||||
180 LuValSArgLeuMetPheArgIleAlaMetArgIleLeuGlyCys 196
|||||
564 TGACCCGACATTAACCTACGGGACGCTTGGG....TCAGCCTGC 607
|||||
197 GluProGlnLeuAlaGlyAspGlyAspSerGluGlnGlnLeuValGluAl 213
|||||
608 GAG.....AAGGACAAGACGATTCATTAACAT 636
|||||
213 aPheGluGlnMetThrArg.....AsnLeuPheSerLeuProIleAsp 228
|||||
637 CTTCACAAAGATCTCTAAGCTTTTGGAGCCTTCAACGTCGCGATTTCA 686
|||||
228 aLProPheSerGlyLeuTyr.....ArgGlyMet 237
|||||
687 TACCATATTTCCGGTGGATCGATCCGCAAGGATTAACAAGCGGCTCGTG 736
|||||
238 LysAlaArgAsn.....LeuIleHisAlaArgIleGluGlnAs 250
|||||
737 AAGCCCGCTATGATCTAGACGGATTTATGACGATTTTCAGTGAACA 786
|||||
250 nILArGAlAlaLysIleCysGlyLeuArgAlaSerGluAlaGly..... 264
|||||
787 TATGAAGAGAAGAG...AATCAAAACGCTGTGATGATGGGATGTG 833
|||||
265 .....GlnGlyCysLysAspAlaLeuGlnLeuIleGlnHisSer 278
|||||
834 TCGATACCGATATGTTGATGATCTCTTGCTTTTACAGTGAAGAGGCC 883
|||||
279 TrpGluArgGlyGluArgLeuAspMetGln..... 288
|||||
884 AAATTAAGTACGAGACAGCGGATCTTCAAAATTCATCAAACTTACCGC 933
|||||
289 ...AlaLeuYsGlnSerSerThrGluLeuLeuPheGlyGlyHisGluT 304
|||||
934 TGACAAATTCAAAGCAATCATCATGACGTTATGTTCGAGGAACGGAA 983
|||||
304 hTThrAlaSerAlaIaThrSerLeuIleThrTyrLeuGlyLeuTyrPro 320
|||||

```

```

984 CGGTAGCGTCGGCGATAGAGTGGCCCTTAACGAGTTATTACGAGGCC 1033
321 HisValLeuGlnLysValArgGluGluLeu...LysSerLysGlyLeuLe 336
|||||
1034 GAGGATCTAAAGCGGTCCACACAGAACTCCGCCGAAGTCGTGGACTT.. 1081
|||||
336 uCysLysSerAsnGlnAspAsnLysLeuAspMetGluIleLeuGluGlnL 353
|||||
1082 .....GACAGACGAGTTGAAAGAACCCACATCGAAGAGT 1115
|||||
353 eulYsTyrIleGlyCysValIleLysGluThrLeuArgLeuAsnProPro 369
|||||
1116 TGACATATCTCAATGACACATCAAAAGAAACCTTAAGATGACCCACCG 1165
|||||
370 ValProGlyGlyPheArgValAlaLeuLysThrPheGluLeuAsnGlyTy 386
|||||
1166 ATCCCTTCCTCCTCCACGAAACCGCGAGGACACTAGATGACAGGTTT 1215
|||||
386 rGlnIleProLysGlyTPrAsnValIleTyrSerIleCysAspThrHisA 403
|||||
1216 CTTCATTCCTCCAAAGAAATCTGCTGTGATGATCAACCGTTTGGCATAGAC 1265
|||||
403 sPValAlaGluIlePheThrAsnLysGluGluPheAsnProAspArghe 419
|||||
1266 GCGACCCCAACTCTTGGACTGACCGGACACGTTTAAGACATGAGGTTT 1315
|||||
420 SerAlaProHisProGluAsp.....AlaSerArgPheSerPheIlePr 434
|||||
1316 TTGGAAACCGGCGTACCGGATTTCAAGGAGACAAATTCGATGATTAAC 1365
|||||
434 oPheGlyGlyGlyLeuArgSerCysValGlyLysGluPheAlaLysIleL 451
|||||
1366 GTTCGGGTGGGGTCGTAGATCGTCCCGGGTATGCCA..... 1402
|||||
451 eulYsIlePheThrValGluLeuAla.....ArgHisCys 463
|||||
1403 ..CTAGGTTATACGCGCTGACTTAGCCGTGCTCATATTAATTATTCG 1450
|||||
464 ...AspTrpGlnLeuLeuAsnGlyProProThrMetLysThrSer 477
|||||
1451 TTCACGTGGAAATTA.....CCTGATGGGATGAAACCAAGT 1486
|||||
seq_name: /cgn2_6/plodata/2/lna/6B_COMB.seq:us-09-103-840A-2
seq_documentation_block:
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103.840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

```

```

alignment_scores:
Quality: 400.00      Length: 486
Ratio: 1.481         Gaps: 21

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2001, 13:25:25 ; Search time 89.59 Seconds
(without alignments)
336.311 Million cell updates/sec

Title: US-09-668-482-32

Sequence: 2612
1 MGFLPALLASALCTFVLPDLL.....PTVYPVNDLPARETFQGD1 497

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

A.Geneseq_0601.*
1: /SIDSL/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SIDSL/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2612	100.0	497	19	Cytochrome P450RA1
2	2612	100.0	497	19	Mouse retinoid met
3	2437	93.3	497	19	Human cytochrome p
4	2437	93.3	497	19	Human retinoid met
5	1734	66.4	492	19	Cytochrome zp450RA
6	1734	66.4	492	19	Zebrafish retinoid
7	1052	40.3	216	21	Human PSEC64 prote
8	474	18.1	468	21	Arabidopsis thalia
9	461	17.6	513	21	Arabidopsis thalia
10	455.5	17.4	461	21	Arabidopsis thalia
11	455.5	17.4	462	21	Arabidopsis thalia

12	455.5	17.4	465	21	AA646491	Arabidopsis thalia
13	450	17.2	472	18	AAW27153	Arabidopsis thalia
14	450	17.2	472	21	AA644571	Arabidopsis thalia
15	450	17.2	472	21	AA645022	Arabidopsis thalia
16	450	17.2	491	21	AA645021	Arabidopsis thalia
17	450	17.2	492	21	AA644570	Arabidopsis thalia
18	445	17.0	444	21	AA644572	Arabidopsis thalia
19	445	17.0	444	21	AA645023	Arabidopsis thalia
20	440.5	16.9	430	21	AA620784	Arabidopsis thalia
21	423.5	16.2	461	21	AA611836	Arabidopsis thalia
22	423.5	16.2	462	21	AA611835	Arabidopsis thalia
23	423.5	16.2	465	21	AA611834	Arabidopsis thalia
24	406	15.5	481	21	AA630049	Arabidopsis thalia
25	406	15.5	489	21	AA630048	Arabidopsis thalia
26	380.5	14.6	433	21	AA630050	Arabidopsis thalia
27	380	14.5	475	21	AA607676	Arabidopsis thalia
28	358.5	13.7	511	20	AA623341	Amino acid sequenc
29	358.5	13.7	511	21	AA619694	A P450-2 protein 1
30	350.5	13.4	513	22	AA631008	Sweetgum confireyl
31	350.5	13.4	513	22	AA648181	Amino acid sequenc
32	347	13.3	388	21	AA620785	B. napus FSH polyP
33	342	13.1	520	22	AA631007	Arabidopsis thalia
34	342	13.1	520	22	AA648179	Amino acid sequenc
35	341	13.1	520	22	AA648180	B. napus FSH polyP
36	337	12.9	520	18	AAW26640	Arabidopsis thalia
37	337	12.9	520	19	AAW40099	Arabidopsis thalia
38	337	12.9	520	21	AA615188	Arabidopsis ferula
39	333.5	12.8	163	21	AA640557	Human ORFX ORF321
40	314.5	12.0	520	20	AA623917	Arabidopsis thalia
41	311.5	11.9	526	20	AAV05675	Maize ferulate-5-h
42	311	11.9	471	21	AA623013	Arabidopsis thalia
43	311	11.9	479	21	AA623012	Arabidopsis thalia
44	308.5	11.8	469	21	AA623014	Arabidopsis thalia
45	300	11.5	576	20	AAV09190	Soybean cytochrome

ALIGNMENTS

RESULT 1	
ID	AAW37735 standard; Protein: 497 AA.
XX	AAW37735;
XX	07-JUL-1998 (first entry)
DE	Cytochrome P450RA1 isoform.
XX	
XX	Retinoid regulated gene; cytochrome P450 gene; enzyme.
KW	oxidative metabolism; P450RA1; retinoid acid; RA; promoter; isoform.
XX	
OS	Synthetic.
XX	
PN	W09749832-AZ.
PD	31-DEC-1997.
XX	
XX	23-JUN-1997; 97WO-CA00488.
PE	
XX	01-OCT-1996; 96US-0724466.
PR	21-JUN-1996; 96US-0667546.
XX	
PA	(TOOH) UNIV QUEENS KINGSTON.
XX	
XX	Petkovich PW;
XX	
XX	WPI; 1998-077193/07.
DR	N-PSDB; AAV09252.
XX	
PT	Identifying DNA encoding inducible or suppressible cytochrome P450 -
PT	by screening for drugs which reduce the catabolism of retinoid acid,
PT	useful in cancer chemotherapy and the treatment of acne and

PT psoriasis
 XX
 PS Disclosure; Pages 59G-59H; 113pp; English.
 XX
 CC This amino acid sequence is of an isoform of cytochrome zp450RA1.
 CC Its expression is dependent on the presence of retinoic acid (RA).
 CC The retinoid-regulated genes such as the inducible cytochrome p450RA1
 CC gene specifically metabolises a derivative of the RA. The cytochrome
 CC p450 gene in general produces enzymes involved in the oxidative
 CC metabolism of endogenous and exogenous compounds. The cytochrome p450
 CC nucleotide sequence can be used to induce or suppress the expression
 CC of its protein. p450RA1 is highly induced by RA in cell lines and
 CC tissues. This allows for development of a drug screen using promoters
 CC and nucleotide sequences to identify drugs which are useful for
 CC reducing the catabolism of RA.
 CC
 XX
 SQ Sequence 497 AA;

Query Match 100.0%; Score 2612; DB 19; Length 497;
 Best Local Similarity 100.0%; Pred. No. 9e-244;
 Matches 497; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGLPALASALCTFVLLFLTAALKIMDLVCVSSRDRSCALPPPGTMGPPFGETLQM 60
 DB 1 mglpallasalctfvpillflaalklwdlycvssdrscalpppgtmgppfgeltqm 60
 QY 61 VLQRRKFLQMKRRKRYGFIYKTHLFGPPYRVKGMADNVRILLGHRILVSVHPASVFTIL 120
 DB 61 vlqrrkflqmkrrkygfiythlfgpprvyvmgadnvrillgehrlysvhwpasvrtll 120
 QY 121 GAGCLSNLHDSHKKORRKVIQAFSREALOCYVLAIEVSSCLEQWLSCGERGLVYPE 180
 DB 121 gagclsnlhdsghkqrkvimgafstrealqcyvliavevssclegwlsogergllvype 180
 QY 181 VKRMFRIMRILLCGEPGAGGEDDQVLEAFEMTRNLFSPLIDVPFSGLYRGVKAR 240
 DB 181 vkrmfrimrilcgepgaggeddqvleafeemtrnlfsplidvpfsglyrgvkar 240
 QY 241 NLHARIEENIRAKIRRIQATEPDGCKDALQTLIEHSMERGEKLDQALQKOSTELLFG 300
 DB 241 nlharieenirakirrligatepdgckdalqtliehsmergerldmqalqksteltllfg 300
 QY 301 GHETTASANSLLTYLGLYPHYLQKVRBEIKSKGLLCKSNQDNKLDMETLEQLKYICV 360
 DB 301 ghettaasanslltylglpvhylqkvreekskglcksnqdnkldmetleqlyicv 360
 QY 361 KETLRMPVPVGGFRRVAKTFPELNGYQIPKGMNITYSICDTHVDADIETKKEFNPDREI 420
 DB 361 ketlrmpvpvgfrrvalktfelngyqipkgnvnyisicdthvdadiftnkeetnpdrtll 420
 QY 421 VPHEDASRSFIFPGGGLRSVCVGEFAKILKIFVELARHCDMDLLNPPWTKTSPTV 480
 DB 421 vphedasrsfifpgggllrsvcvgefakillkifvelarhcdmdllnppwtksptv 480
 QY 481 YPVNDNLPAFFYFGQDI 497
 DB 481 ypvndnlparfitygdi 497

RESULT 2
 AAM44161
 ID AAM44161 standard; Protein; 497 AA.
 XX

AC AAM44161;
 XX
 DT 22-JUN-1998 (first entry)
 XX

DE Mouse retinoid metabolising protein mp450RA1.
 XX
 XX Retinoid metabolising protein; p450RA1; retinoid oxidase;
 KW retinoic acid; mouse; inhibitor; antibody; cancer;
 KW actinic keratosis; oral leukoplakia; head tumour; neck tumour;

KW non-small cell lung carcinoma; basal cell carcinoma;
 KW acute promyelocytic leukaemia; skin cancer; acne; psoriasis;
 KW ichthyosis; therapy; diagnosis; screening.
 XX
 OS Mus musculus.
 XX
 PN W09749815-A1.
 XX
 PD 31-DEC-1997.
 XX
 PF 23-JUN-1997; 97MO-CA00440.
 XX
 PR 01-OCT-1996; 96US-0724466.
 PR 21-JUN-1996; 96US-0667546.
 XX
 PA (TOOH) UNIV QUEENS KINGSTON.
 XX
 PI Beckett BR, Jones G, Petkovich PM, White JA;
 XX
 DR WPI; 1998-077178/07.
 DR N-PSDB; AAV12205.
 XX

PT Retinoid metabolising protein - useful to develop products to treat,
 PT e.g. cancer, actinic keratosis, oral leukoplakia, acne, psoriasis or
 PT ichthyosis
 XX
 PS Claim 1; Page 65-66; 110pp; English.
 XX

CC This protein comprises a novel mouse retinoid metabolising protein,
 CC designated mp450RA1. Its amino acid sequence was deduced from a
 CC cDNA clone (see AAV12205) isolated from a retinoic acid-treated P19
 CC teratocarcinoma library. It includes a haem-binding motif
 CC characteristic of cytochrome p450 proteins. mp450RA1 is a retinoid
 CC oxidase that has the ability to hydroxylate retinoic acid at the 4
 CC position of the beta-ionone ring, and is inducible in epithelial
 CC cells exposed to retinoic acid. Zebrafish, human and mouse p450RA1s
 CC (see AAM44159-61) are claimed. They can be expressed in host cells
 CC and used to metabolise retinoic acid in an organism or cell, in drug
 CC screening, and to raise antibodies useful for inhibiting retinoid
 CC acid hydroxylation for the treatment of cancer, actinic keratosis,
 CC oral leukoplakia, secondary tumours of the head and/or neck,
 CC non-small cell lung carcinomas, basal cell carcinomas, acute
 CC promyelocytic leukaemia, skin cancer, and premalignancy associated
 CC with actinic keratosis, acne, psoriasis and/or ichthyosis.
 CC
 XX
 SQ Sequence 497 AA;

Query Match 100.0%; Score 2612; DB 19; Length 497;
 Best Local Similarity 100.0%; Pred. No. 9e-244;
 Matches 497; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGLPALASALCTFVLLFLTAALKIMDLVCVSSRDRSCALPPPGTMGPPFGETLQM 60
 DB 1 mglpallasalctfvpillflaalklwdlycvssdrscalpppgtmgppfgeltqm 60
 QY 61 VLQRRKFLQMKRRKRYGFIYKTHLFGPPYRVKGMADNVRILLGHRILVSVHPASVFTIL 120
 DB 61 vlqrrkflqmkrrkygfiythlfgpprvyvmgadnvrillgehrlysvhwpasvrtll 120
 QY 121 GAGCLSNLHDSHKKORRKVIQAFSREALOCYVLAIEVSSCLEQWLSCGERGLVYPE 180
 DB 121 gagclsnlhdsghkqrkvimgafstrealqcyvliavevssclegwlsogergllvype 180
 QY 181 VKRMFRIMRILLCGEPGAGGEDDQVLEAFEMTRNLFSPLIDVPFSGLYRGVKAR 240
 DB 181 vkrmfrimrilcgepgaggeddqvleafeemtrnlfsplidvpfsglyrgvkar 240
 QY 241 NLHARIEENIRAKIRRIQATEPDGCKDALQTLIEHSMERGEKLDQALQKOSTELLFG 300
 DB 241 nlharieenirakirrligatepdgckdalqtliehsmergerldmqalqksteltllfg 300
 QY 301 GHETTASANSLLTYLGLYPHYLQKVRBEIKSKGLLCKSNQDNKLDMETLEQLKYICV 360

Db	301	gheftaaatsllfitygllyphvlgkvreeelskxgllcksinqdnkldmetleqikyigcvll	360
QY	361	ketrlrlmnprrpgcgrvalakffelfelngvdiPKGMWNTYSICDTHVAOIFTNKEEENPRFI	420
Db	361	ketrlrlmnprrpgcgrvalakffelfelngvdiPKGMWNTYSICDTHVAOIFTNKEEENPRFI	420
QY	421	vpRHEDASRRSFIFPGGSLGSCVCKEPAKILLIKFTWELRHCDMOLLNPPIMTKTSPTV	480
Db	421	vpRHEDASRRSFIFPGGSLGSCVCKEPAKILLIKFTWELRHCDMOLLNPPIMTKTSPTV	480
QY	481	ypvDNLpARFMYFGODI	497
Db	481	ypvDNLpARFMYFGODI	497

RESULT	3
AAW37734	
ID	AAW37734 standard; Protein; 497 AA

AC	AAW37734;
XX	
DT	07-JUL-1998 (first entry)

Human cytochrome P450RAI protein.

Retinoid regulated gene: cytochrome P450 gene; enzyme:

oxidative metabolism; P450RAI; retinoic acid; RA; promoter. KW

aa Homo sapiens

PN W09749832-A2

PD 31-DEC-1997.

PF 23-JUN-1997;

PR 01-OCT-1996; 96US-0724466

XX

XX

XX

DR N-PSDB; AAV09247.

PT Identifying DNA encoding inducible or suppressible cytochrome P450

PT useful in cancer chemotherapy and the treatment of acne and

XX

AA This is the amino acid sequence of the human cytochrome P450RA1. Its
CC expression is dependent on the presence of retinoic acid (RA). The
CC retinoid-regulated genes such as the inducible cytochrome P450RA1
CC gene specifically metabolises a derivative of the RA. The cytochrome
CC P450 gene in general produces enzymes involved in the oxidative
CC metabolism of endogenous and exogenous compounds. The cytochrome
CC P450 nucleotide sequence can be used to induce or suppress the
CC expression of its protein. P450RA1 is highly induced by RA in cell
CC lines and tissues. This allows for development of a drug screen
CC using promoters and nucleotide sequences to identify drugs which are
CC useful for reducing the catabolism of RA.

.... SQ Sequence 497 AA;

Query Match	93.3%;	Score 2437;	DB 19;	Length 497;
Best Local Similarity	93.4%;	Pred. No. 7.3e-227;		
Matches 464;	Conservative 14;	Mismatches 19;	Indels 0;	Gaps 0;

```

0Y      1 MGLPDLASALCFVFDPLLLFLAALKIMDIYCVSSSDRSCALPLPFGTMGFPPFGTLOM 60
Db      1 mglpdlasalcftvfdplllflaalkimdiycvsssdrcsalplpfgtmgfpfpfgtlo 60
0Y      61 VLORRFLQMRKRRYKGFITYKHTLGFRTVWGMADWVRILLGEHRLVSVHMDASYRITL 120
Db      61 vlorrflqmrkrrygfitykhtlgeftrtvwgmadvvrillgehrlvsvhmdasyritl 120
0Y      121 GAGCLSNHDSHKKORRKYVMQARSRALOCYULVLAEEVSOCLQOMSGEGEGLLYVE 180
Db      121 gagclsnhdshkkrkyvmqarsralocylulvaeevsoclqomsgegeglyllyve 180
0Y      181 VKRLMFRITAMILLKCEGEPGAPGCGEDEQOLEVFEEMTRNLFSLPIIDVPESSLKYGVKAR 240
Db      181 vkrlmfritamillkcegpgapgcgedeqolevfeemtrnlfsldpivepslkygvkar 240
0Y      241 NLIHARIEENIRAKTRIRLQATERPQGGCKDMLQILLIEHSMERGELOMQALKOSTETELTG 300
Db      241 nliharieeniraktrirqlaterpqggckdmlqilliehsmergehqalqostetel 300
0Y      301 GHETASATSLITYLGLYPRVHLQVREELIKSKGLCKSNQDNKLMETLEODIKYICVYI 360
Db      301 ghettasatstlitylglprvhlqvreeleksgllcksnqdnklmetleodikyicvyi 360
0Y      361 KETTLRLNPPVPGGFVYALKTEFLNGYQIPKKNMYTISICQHNVDADLFTKKEFNPDREI 420
Db      361 kettlrlnppvggfvyalktelfngyqipkknmytisicqhnvdadlftkkefnpdre 420
0Y      421 VHPEDASRFSIFPGGGLRSCVCGKEFAKILKFTVELARHNCQWOLNPGPTMKTSPTV 480
Db      421 vhpedasrfsifpggglrscvcgkefakilktfvelarhncqwnlpgptmktsptv 480
0Y      481 YPVDNLPAFTYFGCDI 497
Db      481 ypvndlparftfhgei 497

```

RESULT	4
AAW44160	
ID	AAW44160 standard; Protein; 497 AA

AAW44160; AC

DT 22-JUN-1998 (first entry)

Human retinoid metabolising protein hp450RAI

KW Retinoid metabolising protein; P450RAI; retinoid oxidase;

KW actinic keratosis; oral leukoplakia; head tumour; neck tumour

KW acute promyelocytic leukaemia; skin cancer; acne; psoriasis;

XX

XX

XX

XX

XX

PR 21-JUN-1996; 96US-0667546

PA (TOOH) UNIV QUEENS KINGSTON.

PI Beckett BR, Jones G, Petkovich PM, White JA

DR WPI; 1998-077178/07.

XX defined metabolic protein - useful] to develop products to treat

PT e.g. Cancer, actinic keratosis, oral leukoplakia, acne, psoriasis or
 PT Ichthyosis

PS Claim 1: Page 54-55; 110pp; English.

CC This protein comprises a novel human retinoid metabolizing protein,
 CC designated hp450RAI. Its amino acid sequence was deduced from a
 CC cDNA clone (see AAV12204) isolated from a retinoic acid-treated NR2
 CC cell library. It includes a haem-binding motif characteristic of
 CC cytochrome P450 proteins. hp450RAI is a retinoid oxidase that has
 CC the ability to hydroxylate retinoic acid at the 4 position of the
 CC beta-ionone ring, and is inducible in epithelial cells exposed to
 CC retinoic acid. Zebrafish, human and mouse P450RAIs (see AAW4159-61)
 CC are claimed. They can be expressed in host cells and used to
 CC metabolize retinoic acid in an organism or cell, in drug screening,
 CC and to raise antibodies useful for inhibiting retinoic acid
 CC hydroxylation for the treatment of cancer, actinic keratosis, oral
 CC leukoplakia, secondary tumours of the head and/or neck, non-small
 CC cell lung carcinomas, basal cell carcinomas, acute promyelocytic
 CC leukaemia, skin cancer, and premalignancy associated with actinic
 CC keratosis, acne, psoriasis and/or ichthyosis.

CC Sequence 497 AA.

Query Match 93.3%; Score 2437; DB 19; Length 497;
 Best Local Similarity 93.4%; Pred. No. 7.3e-227;
 Matches 464; Conservative 14; Mismatches 19; Indels 0; Gaps 0;

QY 1 MGPLALASALCTFVLPFLFLTAALKIMDLVCSSDRSCALPLPPTGMPFFGFTLQM 60
 DB 1 MGPLALASALCTFVLPFLFLTAALKIMDLVCSSDRSCALPLPPTGMPFFGFTLQM 60
 QY 61 VIQRKFLQMKRRKYGFIYKTHLFGRPYRVMGADNVRRLIGEHRLVSVHPASVRTIL 120
 DB 61 VIQRKFLQMKRRKYGFIYKTHLFGRPYRVMGADNVRRLIGEHRLVSVHPASVRTIL 120
 QY 121 GAGCISNLDHSHKQKRVKIMQAFSEALQCYVLYAEVSSCLEQWLSGGERGLLYPE 180
 DB 121 GAGCISNLDHSHKQKRVKIMQAFSEALQCYVLYAEVSSCLEQWLSGGERGLLYPE 180
 QY 181 VKRLMFRIMRILLCGCEPPAGGDEQOLVFAFEMETRNLSLPIDVPSGLYRGVKAR 240
 DB 181 VKRLMFRIMRILLCGCEPPAGGDEQOLVFAFEMETRNLSLPIDVPSGLYRGVKAR 240
 QY 241 NLIHRIEENIRAKTRRLQATEPDDGCKDALQLLIHSWGERERLDMQALKOSTELLFG 300
 DB 241 NLIHRIEENIRAKTRRLQATEPDDGCKDALQLLIHSWGERERLDMQALKOSTELLFG 300
 QY 301 GHEETASATSLITVLGYPHYLOKVRREIRKSGILCKSNQDNKIDMETLEOLKTYIGCVI 360
 DB 301 GHEETASATSLITVLGYPHYLOKVRREIRKSGILCKSNQDNKIDMETLEOLKTYIGCVI 360
 QY 361 KETRLNPNVPGGFVNAKTEELANGYQIPKGNVYISICDTHDVADIFNKEEPPNDRFI 420
 DB 361 KETRLNPNVPGGFVNAKTEELANGYQIPKGNVYISICDTHDVADIFNKEEPPNDRFI 420
 QY 421 VPHEDARSRFPIPGGGISCVGKEKFAKILIKRTVELARHQQMQLNGPPTMKTSPTV 480
 DB 421 VPHEDARSRFPIPGGGISCVGKEKFAKILIKRTVELARHQQMQLNGPPTMKTSPTV 480
 QY 481 YPVNLPARFTYFOGDI 497
 DB 481 YPVNLPARFTYFOGDI 497

RESULT 5
 ID AAW37733 standard; Protein: 492 AA.

AC AAW37733;
 XX 07-JUL-1998 (first entry)

XX Cytochrome zp450RAI protein.

XX Retinoid regulated gene; cytochrome P450 gene; enzyme;
 KW oxidative metabolism; P450RAI; retinoic acid; RA; promoter.

XX Danio rerio.

XX WO9749832-A2.

XX 31-DEC-1997.

XX 23-JUN-1997; 97MO-CA00488.

XX 01-OCT-1996; 9605-0724466.

XX 21-JUN-1996; 9605-0667546.

XX (TOOH) UNIT QUEENS KINGSTON.

XX Petkovich PM;

XX WPI; 1998-077193/07.

XX N-PSDB; AAW09251.

PT Identifying DNA encoding inducible or suppressible cytochrome P450 -
 PT by screening for drugs which reduce the catabolism of retinoic acid,
 PT useful in cancer chemotherapy and the treatment of acne and
 PT psoriasis

PS Example 1; Pages 53-55; 113pp; English.

CC This is the amino acid for cytochrome zp450RAI of the zebra fish.
 CC Its expression is dependent on the presence of retinoic acid (RA).
 CC The retinoid-regulated genes such as the inducible cytochrome P450RAI
 CC gene specifically metabolises a derivative of the RA. The cytochrome
 CC P450 gene in general produces enzymes involved in the oxidative
 CC metabolism of endogenous and exogenous compounds. The cytochrome
 CC P450 nucleotide sequence can be used to induce or suppress the
 CC expression of its protein. P450RAI is highly induced by RA in cell
 CC lines and tissues. This allows for development of a drug screen using
 CC promoters and nucleotide sequences to identify drugs which are useful
 CC for reducing the catabolism of RA.

CC Sequence 492 AA;

Query Match 66.4%; Score 1734; DB 19; Length 492;
 Best Local Similarity 67.7%; Pred. No. 6e-159;
 Matches 336; Conservative 61; Mismatches 89; Indels 10; Gaps 4;

QY 1 MGPLALASALCTFVLPFLFLTAALKIMDLVCSSDRSCALPLPPTGMPFFGFTLQM 60
 DB 1 MGPLALASALCTFVLPFLFLTAALKIMDLVCSSDRSCALPLPPTGMPFFGFTLQM 60
 QY 61 VIQRKFLQMKRRKYGFIYKTHLFGRPYRVMGADNVRRLIGEHRLVSVHPASVRTIL 120
 DB 61 VIQRKFLQMKRRKYGFIYKTHLFGRPYRVMGADNVRRLIGEHRLVSVHPASVRTIL 120
 QY 121 GAGCISNLDHSHKQKRVKIMQAFSEALQCYVLYAEVSSCLEQWLSGGERGLLY 177
 DB 121 GAGCISNLDHSHKQKRVKIMQAFSEALQCYVLYAEVSSCLEQWLSGGERGLLY 177
 QY 178 YPEVRLMFRIMRILLCGCEPPAGGDEQOLVFAFEMETRNLSLPIDVPSGLYRGV 237
 DB 178 YPEVRLMFRIMRILLCGCEPPAGGDEQOLVFAFEMETRNLSLPIDVPSGLYRGV 237
 QY 238 KANLIHRIEENIRAKTRRLQATEPDDGCKDALQLLIHSWGERERLDMQALKOSTEL 297
 DB 238 KANLIHRIEENIRAKTRRLQATEPDDGCKDALQLLIHSWGERERLDMQALKOSTEL 297
 QY 298 LEGETTASATSLITVLGYPHYLOKVRREIRKSGILCKSNQDNKIDMETLEOLKTYIG 357
 DB 298 LEGETTASATSLITVLGYPHYLOKVRREIRKSGILCKSNQDNKIDMETLEOLKTYIG 357

QY 358 CVIETHETRLANPPYPGGRVALTKFEELNGVOIPKGMWVIVISICDFHVAOIFPNKKEEFPND 417

Db 354 cvkeltlirpappvggfrvalkltfelngyqipkwmvnylsicdthdvadvipnkeefpe 413

QY 418 RFYIPHEDEASRFSFIPFGGILRSCVCKEFAKILKIFVLELARHCDWOLNGPPTMTS 477

Db 414 rfmkgjleodsrfirnyipfggsrmcwgketakvllkflvltqhnwllnsgpptmtkg 473

QY 478 PTVYVDNLPARETFYE 493

Db 474 ptiypvdnlplekftsyz 489

RESULT 6

AAW44159

AAW44159 standard; Protein; 492 AA.

AC AAW44159;

XX 22-JUN-1998 (first entry)

DE Zebrafish retinoid metabolising protein zp450RA1.

XX Retinoid metabolising protein; P450RA1; retinoid oxidase;

XX retinoic acid; zebrafish; inhibitor; antibody; cancer;

XX actinic keratosis; oral leukoplakia; head tumour; neck tumour;

XX non-small cell lung carcinoma; basal cell carcinoma;

XX acute promyelocytic leukaemia; skin cancer; acne; psoriasis;

XX ichthyosis; therapy; diagnosis; screening.

OS Danio rerio.

XX MO9749815-A1.

PN 31-DEC-1997.

PD 23-JUN-1997; 97WO-CA00440.

PF 01-OCT-1996; 96US-0724466.

PR 21-JUN-1996; 96US-0667546.

XX (TOOH) UNIV QUEENS KINGSTON.

PA Beckett BR, Jones G, Petkovich PM, White JA;

XX WPI, 1998-077178/07.

DR N-PSDB; AAV12203.

XX Retinoid metabolising protein - useful to develop products to treat,

PT e.g. cancer, actinic keratosis, oral leukoplakia, acne, psoriasis or

PT ichthyosis

XX Claim 1; Page 50-51; 110pp; English.

XX This protein comprises a novel zebrafish retinoid metabolising

XX protein, designated zp450RA1. Its amino acid sequence was deduced

XX from a cDNA clone (see AAV12203) isolated from a 6-18 hr embryo

XX library. It includes a haem-binding motif characteristic of

XX cytochrome P450 proteins. zp450RA1 is a retinoid oxidase that has

XX the ability to hydroxylate retinoic acid at the 4 position of the

XX beta-ionone ring, and is inducible in epithelial cells exposed to

XX retinoic acid. Zebrafish, human and mouse P450RA1s (see AAW44159-61)

XX are claimed. They can be expressed in host cells and used to

XX metabolize retinoic acid in an organism or cell, in drug screening,

XX and to raise antibodies useful for inhibiting retinoic acid

XX hydroxylation for the treatment of cancer, actinic keratosis, oral

XX leukoplakia, secondary tumours of the head and/or neck, non-small

XX cell lung carcinomas, basal cell carcinomas, acute promyelocytic

XX leukaemia, skin cancer, and premalignancy associated with actinic

XX keratosis, acne, psoriasis and/or ichthyosis.

XX Sequence 492 AA.

[illegible]

[illegible]

PD 17-AUG-2000.
XX
XX 11-FEB-2000; 2000MO-US03820.
XX
PR 11-FEB-1999; 99US-0119657.
PR 11-FEB-1999; 99US-0119658.
XX
XX (ARIZ-) ARIZONA BOARD OF REGENTS.
PI Aspiroz R, Choe S, Feldmann KA.
XX WPI; 2000-549142/50.
DR N-PSDB; AAA59399.
XX
XX New isolated dwf4 polynucleotide useful for altering the phenotype of
plants, for diagnostic assays and in the production of antibodies -
PS Claim 50; Fig 11; 113pp; English.
XX
XX The present sequence represents a DMF4 polypeptide. The polypeptide is a
CC cytochrome P450 enzyme that mediates multiple steps in synthesis of
CC brassinosteroids. Specifically, it mediates multiple
CC 22alpha-hydroxylation steps in brassinosteroid biosynthesis. The DMF4
CC polynucleotide is used for altering the phenotype of a plant. DMF4
CC plants display a dramatic reduction in the length of different organs,
CC and this size reduction is attributable to a defect in cell elongation.
CC The DMF4 polynucleotides and polypeptides can be used in diagnostic
CC assays and to generate antibodies, which can be used to produce
CC immunogenic compositions.
XX
XX Sequence 513 AA;
SQ

Query Match 17.6%; Score 461; DB 21; Length 513;
Best Local Similarity 28.4%; Pred. No. 6.2e-36;
Matches 148; Conservative 78; Mismatches 231; Indels 64; Gaps 14;

QY 3 LPAALLSALCTFVPLLLTLAALKMDLYCVSHRSCALPLPGTGMFFFGETL----58
DB 10 Lp11111ps1-----1s1111111k-----rnrktrfnlppgksgwplgetliylk 57
QY 59 -QNVLDQRKRLQMKRRKRYGFIYKTHLFGRETVRMGADNVRRIILGSHRLVSNVWPSVR 117
DB 58 pyatctlgdmqghvsygyklyrtnlfgreplvsadaglnrffllqnegrlfecyspsig 117
QY 118 TLGAGCLSMHDSHKKORRVIMQAFSRALOCYVLV-IAEVSVCLEQWLSCGERGL 176
DB 118 gllgkwsmlvlgvdhmrsislnflsharltlllkdrerhltfvidswqg--nsifs 175
QY 177 VYEVVKRMFRIMRIILGCEPRPAGGDEQOLVEAFEEKTRNLFSPLDVPFSGLYRG 236
DB 176 agdeakktfnlnakhlmsndpge---eeteqkkeyvlfmkyvnapnlpjstayhka 231
QY 237 VKARNLIHARIEENIRAKIRLQATEPDDGCKDALDLLEHSWE-----RGRRLDMQ- 288
DB 232 lgratallkfkierkme--tkldikeed---geeevkteeaeemsksdhvtqtddd1 286
QY 289 ---ALKQS---STP-----LLFGSHETTSATSLITVGLYPRVYLQKREIKSGKL 335
DB 287 lgvvlkhnslstegldldl1sllfaghetssvalaiafflgacpkaeeleehleiar 346
QY 336 LCKSNDNKLDMETLSQLKTYIGCVIKETILRLNPPVGGFVALKTEELNGYQIRPKGNVT 395
DB 347 akkelgeselnwddykkmftfcvlnetlrlgnvrvflhkaikdyrykysgkwkvl 406
QY 396 YSICDTHVDADITTNKEEPNDRFIVRHPEDASRFS-----FLPRGGGLRSCVCKGE 446
DB 407 pvlsevhlnsrydgnrlfnrvwtgqmgasssgsfsftwgnmymprjg9pdrclagse 466
QY 447 FAKILIKITFVELARHCQWOLNGRPRTMKTSPVYVVDNLP 487
DB 467 laktlemavfihlvlkfnwelae-----ddqprafvfvdtr 502

RESULT 10
AAG46490
ID AAG46490 standard; Protein: 461 AA.
XX
XX AAG46490;
AC
XX
XX 18-OCT-2000 (first entry)
DT
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 58494.
DE
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX
XX Arabidopsis thaliana.
OS
XX
XX EP1033405-A2.
PN
XX
XX 06-SEP-2000.
PD
XX
XX 25-FEB-2000; 2000EP-0301439.
PF
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.

```
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 02-JUL-1999; 99US-0142154.  
PR 06-JUL-1999; 99US-0142380.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 21-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-01452913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 03-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.
```

```
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151338.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154479.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.
```

Query Match 17.48; Score 455.5; DB 21; Length 461;
Best Local Similarity 28.28; Pred. No. 1.8e-35;
Matches 137; Conservative 91; Mismatches 211; Indels 47; Gaps 13;

```
QY 1 MGRLALASALCTFVLLPLLEFLALTKLMDLYCVSSRRSCALPLPPTGMPFPGETIOM 60  
DB 4 mglillilivs-lc-----salllr-----ngmrytknglppptumwplfigetlef 46  
QY 61 VLORRKFLOMRKRYGFIYKTHLFGRPYRVMGADNVRRILLGHRILVSVMWPAVSFTIL 120  
DB 47 lkqpnfmrngrlrygstfkehllgcptllismdeevnryllkneskylvpypqsmidl 106  
QY 121 GAGCISNLHDSHKKORKKVTIMQAFSRALQCYVIVIAEF-VSSGLRQWMLSGERGILL-VY 178  
DB 107 gtcmaavhgsshrllmrgsllslsttmtrchlllpkvdhmrtyldqg---nelevldiq 163
```


PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148365.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0156559.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158339.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.

PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 17.4%; Score 455.5; DB 21; Length 462;
Best Local Similarity 28.2%; Pred. No. 1,8e-35;
Matches 137; Conservative 91; Mismatches 211; Indels 47; Gaps 13;

QY 1 MGLPALLASALCTFVLPULFLFAAKIMPDYCVSSDRSCALPLPGTGPFPFGTLOM 60
DB 5 mglillllivs-lc-----sallrw-----nmrytknglppgtunwplfgettef 47
QY 61 VLORRRFLQMKRRKRGFYKTHLEFGPTVVRMGADNVRRILGHNRLVSVHPASVYRTIL 120
DB 48 lkqgnfmtrngtrlygsffkshllgcpptlismdsevnryllkneskglvpgypqsmldll 107
QY 121 GAGCLSNLHDSHQRKRVIMQAFSREALQCYLVIAEP-VSSCLQOMLSCGERGLL-VY 178
DB 108 gtcomaavbgshrlmrgslslstsmmrchllpkvchfmrsgldqw--nelevyldq 164
QY 179 PEVKRLMFRIMARLLLCGCEPGPAGGDEBOQVLEAFPEMRNLFSPLIDVPFSGLYRGVK 238
DB 165 dktkhmaflsslqtagnlrkpf----veefktafkllvgvgtlsvpidpgtunrcgldq 219
QY 239 ARNLIHARIEENIRAKRRRLQATEPPDG--CKDALQLLIEHSWGERRLDMQALKOSTE 296
DB 220 arn-----nidrllrlrelmgerdsgetfdmlgylmkkgunrypltd-eeirdqvt 270
QY 297 LLEFGHETTASATSLITFYGLVPHVLQKVBREIKSGLLCKSNQDNKLDMELEQLKYI 356
DB 271 llysgyetvstsmmalakyldhdpkqlqelrae--hlaferkrqdeplgyledevksmkft 328
QY 357 GCYVKEFTLRNLPPVPGGFRVALKTFEELNGVQIPKGMWVYSICDTHVDADIFTNKEEFP 416
DB 329 ravigetstiaivngylrkttrtleinsgylipkwrllyytreindandlyedpllnp 388
QY 417 DRFTVPHEDASRFSFIIPFGGSLRSCVGEKFAKILKIFVEALARHCDQQLNGPPTMKT 476
DB 389 wrwmkksle--sgnscfvgggttlcpkqelgveisflhyftryrweeigddelm-- 444
QY 477 SPVYYP 482
DB 445 --vfp 447

RESULT 12
AAG46491 standard; Protein; 465 AA.
ID AAG46491

AC AAG46491;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 58495.

XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

OS Arabidopsis thaliana.

PN EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX

PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0123788.
PR 23-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132464.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0138847.
PR 16-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.

PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 20-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 22-AUG-1999; 99US-0149902.
PR 22-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.

```

PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159284.
PR 13-OCT-1999; 99US-0159285.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160960.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

Query Match 17.4%; Score 455.5; DB 21; Length 465;
 Best Local Similarity 28.2%; Pred. No. 1.8e-35;
 Matches 137; Conservative 91; Mismatches 211; Indels 47; Gaps 13;

```

QY 1 MGLPALLSAICTFVPLPLFLAALAKLMDLYCVSSRDBSCALPRLPGTGMGPFPGETLQM 60
DB 8 mgl111111vs-1c-----sall1w-----nqm1ytkngl1ppgtm1wpl1getlef 50
QY VLQRRKFLQMKRRKRYGTYKTHLFGRTVAVMGADNVRRLILGEHRLVSVHMPASVRTL 120
DB 51 lkqgnfntnqrllygsffkshlllgcprtlismdeevnyllkneksqlypypqsmldil 110
QY 121 GAGCLSNLHDSHQRKQKVVIMQAFSREALQCYVLYIAEE-VSSCLEQWLSGGERGL-VY 178
DB 111 gtcmavaahvgshtlmrgsl1s1stmmrdhllpkvdhfmrsyldqv---nelevldiq 167
QY 179 PEVKMLMRIMARILLGCEPBPAGGEDEQQLVEAFEEMTNLSLPLDVPFSGILYRKYV 238
DB 168 dktkmaflssltq1ag1n1rpf-----veefktaffk1lvgl1svpldl1pqt1nrc91q 222
QY 239 ARNL1HARTEENIRAKIRLQATEPDDG--CKDALQLLIEHSWGERGLDQALQKOSTE 296
DB 223 an-----nid1lllelmgerrdsgettdm1gylmkkegn1rpltd-eel1rdqvt 273
QY 297 LIFGHEHTASATSLTYLGLYPRVLOKVREELKSKGLCKSNODNKLDMETLEQLKYI 356
DB 274 llysgyelvstsmmal1ylnhpkalqelrae--hlafrerkrqdeplg1edvksmkft 331
QY 357 GCYIETELANDPNVPGGRVALKTFELNGYOLPKGMWNTYISCTDHDVAD1FTKKEENP 416
DB 332 ravlyetsral1vngv1rkt1rdle1ngyl1pkqwt1yvv1r1e1ngdan1yep1l1np 391
QY 417 DGEIYRHPEDASRFSFIRPGGSLRSCVGEFAKILK1FTVELAHNCOMOLNCPRTKT 476
DB 392 wwmkks1e--sgnsc1fv1g9g1r1cp3k1g1v1e1ss1flly1vtr1ywe1g1delm-- 447
QY 477 SPTVYR 482
1:1

```

```

DB 448 ---vfp 450

RESULT 13
AAW27153
ID AAW27153 standard; Protein; 472 AA.
XX
AC AAW27153;
XX
DT 14-APR-1998 (first entry)
XX
DE Arabidopsis thaliana cytochrome P450-type hydroxylase.
XX
KW Cytochrome P450-type hydroxylase; identification; brassinosteroid;
KW brassinosteroid inhibitor; modified plant; recombinant production;
KW testosterone.
XX
OS Arabidopsis thaliana.
XX
PN MO9735986-A1.
XX
PD 02-OCT-1997.
XX
PF 27-MAR-1997; 97WO-EP01586.
XX
PR 27-MAR-1996; 96US-0622166.
XX
PA (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX
PI Altmann T, Koncz C, Mathur J, Szekeres MA;
XX
DR WPI: 1997-489649/45.
XX
DR N-PDSB; AAT85306, AAT85307.
XX
PT New isolated plant cytochrome P450-type hydroxylase gene - used to
PT identify substances acting as brassino-steroid(s) or brassinosteroid
PT inhibitors for the production of modified plants
XX
PS Claim 1; Pages 44-46; 77pp; English.
XX
XX
CC The present sequence is Arabidopsis thaliana cytochrome
CC P450-type hydroxylase. The hydroxylase can be used to identify
CC brassinosteroids or brassinosteroid inhibitors, useful to produce
CC plants with modified physiological and/or phenotypic
CC characteristics. The modified plants may show, e.g. stimulation of
CC growth, increased cell elongation, increased wood production,
CC accelerated seed germination at low temperatures, an increase in
CC dry weight, repressed anthocyanin production during growth in light
CC and/or inhibited de-etiolation which is induced, e.g. by cytokinln,
CC in the dark or an increase in stress tolerance. The hydroxylase or
CC its coding sequence can also be used for the recombinant production
CC of compounds, e.g. testosterone.
XX
SQ Sequence 472 AA;

```

Query Match 17.2%; Score 450; DB 18; Length 472;
 Best Local Similarity 26.6%; Pred. No. 6.3e-35;
 Matches 134; Conservative 90; Mismatches 217; Indels 62; Gaps 15;

```

QY 6 L1ASALCTFVPLPLFLAALAKLMDLYCVSSRDBSCALPRLPGTGMGPFPGETLQM----- 61
DB 9 l1s1aagf1111-----rrtyr1mg1ppgs1g1p1l1getf1g1ayk 52
QY 62 -LQRRKFLQMKRRKRYGTYKTHLFGRTVAVMGADNVRRLILGEHRLVSVHMPASVRTL 120
DB 53 tenpepfidervarygs1vnt1h1l1fgepl1fsadpetrr1v1lqnegk1fecsypas1cn1l 112
QY 121 GAGCLSNLHDSHQRKQKVVIMQAFSREALQCYVLY-IAEYSSCLEQWLSGGERGLLYVP 179
DB 113 gkhs1llmkgs1hkr1mbs1lms1ans1ikdh1m1d1dr1v1r1n1dws1-----ry1lme 168
QY 180 EVKRLMFRIMARILLGCEPBPAGGEDEQQLVEAFEEMTNLSLPLDVPFSGILYR-GVK 238

```

Db 169 eakktfelvtvkqmsfdp-----gweaseslrkeyllviegffsfppl-!stvyrkaiq 222
QY 239 ARNLHARIEENIRAKIRLQATEPDDG--CKDALQLLI--EHSWERGEERLDMQALKOSS 294
Db 223 arr-----kvaaalvvyvkmkrreeeegearkkdhlaalaaddgfsdeelvdf-----1 272
QY 295 TELFGHEFTFASATSLITVGLYPHVLOKREBIRKSGLLCSNQONKIDMETLEOLK 354
Db 273 vallvayettstmltavkfilteplalaqlkee--hekiramksdyslswgdykmp 330
QY 355 YIGCVIKETRLNLPVPGCFRVALKTEFLNGYQIPKGMNVYSICDTHVADIFTNKEEF 414
Db 331 ftgcgvnetlrvanligvfframtdveikgkikipgkwkfsfravhldpnhfkdartf 390
QY 415 NPDRF----IYHPHEDASRFESFIPIFGGGLRSCVCGKEFAKILKFTVYELARHCWOLLNG 470
Db 391 npwtwgsnsvtctgsnv----fcpfggprlcpgyelarvalsvflhrlvlgfsw----- 441
QY 471 PPTMKTSPTVYVPVNDLPARFTYF 493
Db 442 vpaegdklvffprttrtqkrypif 464

RESULT 14

AAG44571 standard; Protein; 472 AA.

AAG44571.

18-OCT-2000 (first entry)

Arabidopsis thaliana protein fragment SEQ ID NO: 55847.

Protein identification: signal transduction pathway; metabolic pathway;
hybridisation assay; genetic mapping; gene expression control; promoter;
termination sequence.

Arabidopsis thaliana.

EPI033405-A2.

06-SEP-2000.

25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142803.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145214.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145813.
PR 27-JUL-1999; 99US-0145918.

PR	27-JUL-1999	9905-0145919
PR	28-JUL-1999	9905-0145951
PR	02-AUG-1999	9905-0145386
PR	02-AUG-1999	9905-0145388
PR	02-AUG-1999	9905-0145389
PR	03-AUG-1999	9905-0147038
PR	04-AUG-1999	9905-0147204
PR	04-AUG-1999	9905-0147302
PR	05-AUG-1999	9905-0147199
PR	05-AUG-1999	9905-0147260
PR	06-AUG-1999	9905-0147303
PR	06-AUG-1999	9905-0147416
PR	09-AUG-1999	9905-0147493
PR	09-AUG-1999	9905-0147935
PR	10-AUG-1999	9905-0148171
PR	11-AUG-1999	9905-0148319
PR	13-AUG-1999	9905-0148341
PR	13-AUG-1999	9905-0148565
PR	16-AUG-1999	9905-0148684
PR	17-AUG-1999	9905-0149368
PR	18-AUG-1999	9905-0149175
PR	20-AUG-1999	9905-0149426
PR	20-AUG-1999	9905-0149723
PR	20-AUG-1999	9905-0149929
PR	22-AUG-1999	9905-0149902
PR	23-AUG-1999	9905-0149930
PR	25-AUG-1999	9905-0150566
PR	26-AUG-1999	9905-0150884
PR	27-AUG-1999	9905-0151065
PR	27-AUG-1999	9905-0151066
PR	30-AUG-1999	9905-0151103
PR	30-AUG-1999	9905-0151303
PR	15-SEP-1999	9905-0154018
PR	16-SEP-1999	9905-0154019
PR	20-SEP-1999	9905-0154179
PR	20-SEP-1999	9905-0155139
PR	22-SEP-1999	9905-0155486
PR	24-SEP-1999	9905-0156559
PR	28-SEP-1999	9905-0156458
PR	29-SEP-1999	9905-0156596
PR	04-OCT-1999	9905-0157117
PR	05-OCT-1999	9905-0157753
PR	06-OCT-1999	9905-0157865
PR	07-OCT-1999	9905-0158029
PR	08-OCT-1999	9905-0158232
PR	12-OCT-1999	9905-0158369
PR	13-OCT-1999	9905-0159293
PR	13-OCT-1999	9905-0159294
PR	14-OCT-1999	9905-0159310
PR	14-OCT-1999	9905-0159331
PR	14-OCT-1999	9905-0159637
PR	18-OCT-1999	9905-0159638
PR	18-OCT-1999	9905-0159584
PR	21-OCT-1999	9905-0160741
PR	21-OCT-1999	9905-0160767
PR	21-OCT-1999	9905-0160768
PR	21-OCT-1999	9905-0160770
PR	21-OCT-1999	9905-0160814
PR	22-OCT-1999	9905-0160815
PR	22-OCT-1999	9905-0160980
PR	22-OCT-1999	9905-0160981
PR	22-OCT-1999	9905-0161404
PR	25-OCT-1999	9905-0161405
PR	25-OCT-1999	9905-0161406

	Query Match	17.2%	Score 450;	DB 21;	Length 472;
	Best Local Similarity	26.6%	Pred. No. 6,3e-35;		
	Matches 134;	Conservative 90;	Mismatches 217;	Indels 62;	Gaps 1
QY	6	LIASALCTFVLPILLFLFALAKLMDLYCVSSRDSICALPRPGTMGFFPGETLQW----	61		
Db	9	llsisaagfilll-----rrtryrmgippslglpigtetqlgayk	52		
QY	62	-LQRRKFLQMKRRKRGYFITYKTHLFGRTVAVMGADNVRRLIGENHRLVSVHMFASVRTLL	120		
Db	53	tenpepitdevarygsvfmfhligeptfisaodecmrfvigneqklfecsyasiscnll	112		
QY	121	GAGCLSNLHDSHKORKKVKVIMQAFSRALOCYVLV-TAEVSSCLQWMLSCGERGLLYP	179		
Db	113	gkhslllmkgsltkhrmnlmsfnsnsilkhnlmldidrlvrlndlsws-----rvllme	168		
QY	180	EVRKLMEFIAMRILLGCEPPGAGGEDEQQLVEAFEEMTNLSLPIDIVFSGLYR-GVK	238		
Db	169	eakktfcltvlvkgqlmsfdr-----gwseslrkeylllviegflslpirl-fstlyrkaig	222		
QY	239	ARNLIHAREENINAKIRRLQATPEPDG--CKDALQLLI--EHSWGERGERLDMQALKOSS	294		
Db	223	arr-----kvaaelvyywmktrreeeegaerkkmlaalaaadgfsadeelvydf-----1	272		
QY	295	TELLFGHETTASATSLITYLGLYPHYLVQKREELKSGKLLCSNODNKLDMETLEOLK	354		
Db	273	vallvagyetstlmtlavkfltetplalaqlkee--hekiramksdsyslewsdykmp	330		
QY	355	YIGCYIKRTLLNPPVPGCFRVALKTEFLNGYOLPRGMNVYISICDTHVDADIFTNKEEF	414		
Db	331	ftcgcvneltlvaniigvtrramcdvelkygktpkgwkvfssfravhdpnhfkartf	390		
QY	415	NPDRF-----IVPHBEDASRFSFIPFGGGLRSCVGRKEFAKILKIFTEYLARHCDWOLLNG	470		
Db	391	npwrgnsvetlgsnv-----ftpfgggprlcpdyelaravalvflhrlyltgfsaw-----	441		
QY	471	PPTMKTSPTVYPVNDLPRARETYF	493		
Db	442	vpaeqdklvpftrltqkryplf	464		
RESULT 15					
AAG45022					
ID	AAG45022 standard; Protein: 472 AA.				
XX	AAG45022;				
XX	18-OCT-2000 (first entry)				
XX	Arabidopsis thaliana protein fragment SEQ ID NO: 56469.				
XX	Protein identification; signal transduction pathway; metabolic pathway;				
XX	hybridisation assay; genetic mapping; gene expression control; promoter				
XX	termination sequence.				
XX	Arabidopsis thaliana.				
XX	EP1033405-A2.				
XX	06-SEP-2000.				
XX	25-FEB-2000; 2000EP-0301439.				
XX	25-FEB-1999; 99US-0121825.				
PR	05-MAR-1999; 99US-0123180.				

PR 09-MAR-1999; 9905-0123548.
PR 23-MAR-1999; 9905-0125788.
PR 25-MAR-1999; 9905-0126264.
PR 29-MAR-1999; 9905-0126785.
PR 01-APR-1999; 9905-0127462.
PR 06-APR-1999; 9905-0128234.
PR 08-APR-1999; 9905-0128714.
PR 16-APR-1999; 9905-0129845.
PR 19-APR-1999; 9905-0130077.
PR 21-APR-1999; 9905-0130449.
PR 23-APR-1999; 9905-0130510.
PR 28-APR-1999; 9905-0130891.
PR 30-APR-1999; 9905-0131449.
PR 30-APR-1999; 9905-0132048.
PR 04-MAY-1999; 9905-0132407.
PR 05-MAY-1999; 9905-0132484.
PR 06-MAY-1999; 9905-0132485.
PR 06-MAY-1999; 9905-0132486.
PR 07-MAY-1999; 9905-0132487.
PR 11-MAY-1999; 9905-0134256.
PR 14-MAY-1999; 9905-0134218.
PR 14-MAY-1999; 9905-0134219.
PR 14-MAY-1999; 9905-0134221.
PR 14-MAY-1999; 9905-0134370.
PR 18-MAY-1999; 9905-0134376.
PR 19-MAY-1999; 9905-0134941.
PR 20-MAY-1999; 9905-0135124.
PR 21-MAY-1999; 9905-0135353.
PR 24-MAY-1999; 9905-0135629.
PR 25-MAY-1999; 9905-0136021.
PR 25-MAY-1999; 9905-0136392.
PR 28-MAY-1999; 9905-0136782.
PR 01-JUN-1999; 9905-0137222.
PR 03-JUN-1999; 9905-0137528.
PR 04-JUN-1999; 9905-0137724.
PR 07-JUN-1999; 9905-0137724.
PR 08-JUN-1999; 9905-0138094.
PR 10-JUN-1999; 9905-0138540.
PR 10-JUN-1999; 9905-0138847.
PR 14-JUN-1999; 9905-0139119.
PR 16-JUN-1999; 9905-0139452.
PR 16-JUN-1999; 9905-0139453.
PR 17-JUN-1999; 9905-0139459.
PR 18-JUN-1999; 9905-0139492.
PR 18-JUN-1999; 9905-0139454.
PR 18-JUN-1999; 9905-0139455.
PR 18-JUN-1999; 9905-0139456.
PR 18-JUN-1999; 9905-0139457.
PR 18-JUN-1999; 9905-0139458.
PR 18-JUN-1999; 9905-0139459.
PR 18-JUN-1999; 9905-0139460.
PR 18-JUN-1999; 9905-0139461.
PR 18-JUN-1999; 9905-0139462.
PR 18-JUN-1999; 9905-0139463.
PR 18-JUN-1999; 9905-0139464.
PR 18-JUN-1999; 9905-0139750.
PR 21-JUN-1999; 9905-0139763.
PR 22-JUN-1999; 9905-0139817.
PR 23-JUN-1999; 9905-0139899.
PR 23-JUN-1999; 9905-0140353.
PR 23-JUN-1999; 9905-0140354.
PR 24-JUN-1999; 9905-0140695.
PR 28-JUN-1999; 9905-0140823.
PR 29-JUN-1999; 9905-0140991.
PR 30-JUN-1999; 9905-0141287.
PR 01-JUL-1999; 9905-0141842.
PR 01-JUL-1999; 9905-0142154.
PR 02-JUL-1999; 9905-0142055.
PR 06-JUL-1999; 9905-0142390.
PR 08-JUL-1999; 9905-0142803.
PR 09-JUL-1999; 9905-0142920.
PR 12-JUL-1999; 9905-0142977.
PR 13-JUL-1999; 9905-0143542.
PR 14-JUL-1999; 9905-0143624.

PR 15-JUL-1999; 9905-0144005.
PR 16-JUL-1999; 9905-0144085.
PR 16-JUL-1999; 9905-0144086.
PR 19-JUL-1999; 9905-0144325.
PR 19-JUL-1999; 9905-0144331.
PR 19-JUL-1999; 9905-0144332.
PR 19-JUL-1999; 9905-0144333.
PR 19-JUL-1999; 9905-0144334.
PR 19-JUL-1999; 9905-0144335.
PR 20-JUL-1999; 9905-0144352.
PR 20-JUL-1999; 9905-0144632.
PR 21-JUL-1999; 9905-0144884.
PR 21-JUL-1999; 9905-0145086.
PR 21-JUL-1999; 9905-0145088.
PR 22-JUL-1999; 9905-0145085.
PR 22-JUL-1999; 9905-0145087.
PR 22-JUL-1999; 9905-0145089.
PR 22-JUL-1999; 9905-0145192.
PR 23-JUL-1999; 9905-0145145.
PR 23-JUL-1999; 9905-0145218.
PR 23-JUL-1999; 9905-0145224.
PR 26-JUL-1999; 9905-0145276.
PR 27-JUL-1999; 9905-0145913.
PR 27-JUL-1999; 9905-0145918.
PR 27-JUL-1999; 9905-0145919.
PR 28-JUL-1999; 9905-0145951.
PR 02-AUG-1999; 9905-0146386.
PR 02-AUG-1999; 9905-0146388.
PR 02-AUG-1999; 9905-0146389.
PR 03-AUG-1999; 9905-0147208.
PR 04-AUG-1999; 9905-0147204.
PR 04-AUG-1999; 9905-0147302.
PR 05-AUG-1999; 9905-0147192.
PR 05-AUG-1999; 9905-0147260.
PR 06-AUG-1999; 9905-0147303.
PR 06-AUG-1999; 9905-0147416.
PR 09-AUG-1999; 9905-0147493.
PR 09-AUG-1999; 9905-0147935.
PR 10-AUG-1999; 9905-0148171.
PR 11-AUG-1999; 9905-0148319.
PR 12-AUG-1999; 9905-0148341.
PR 13-AUG-1999; 9905-0148565.
PR 13-AUG-1999; 9905-0148684.
PR 16-AUG-1999; 9905-0149368.
PR 17-AUG-1999; 9905-0149175.
PR 18-AUG-1999; 9905-0149426.
PR 20-AUG-1999; 9905-0149722.
PR 20-AUG-1999; 9905-0149723.
PR 20-AUG-1999; 9905-0149929.
PR 23-AUG-1999; 9905-0149902.
PR 23-AUG-1999; 9905-0149930.
PR 25-AUG-1999; 9905-0150566.
PR 26-AUG-1999; 9905-0150884.
PR 27-AUG-1999; 9905-0151065.
PR 27-AUG-1999; 9905-0151066.
PR 27-AUG-1999; 9905-0151080.
PR 30-AUG-1999; 9905-0151303.
PR 31-AUG-1999; 9905-015138.
PR 01-SEP-1999; 9905-0151930.
PR 07-SEP-1999; 9905-0152363.
PR 10-SEP-1999; 9905-0153070.
PR 13-SEP-1999; 9905-0153758.
PR 15-SEP-1999; 9905-0154018.
PR 16-SEP-1999; 9905-0154039.
PR 20-SEP-1999; 9905-0154779.
PR 22-SEP-1999; 9905-0155139.
PR 23-SEP-1999; 9905-0155486.
PR 24-SEP-1999; 9905-0155659.
PR 28-SEP-1999; 9905-0156458.
PR 29-SEP-1999; 9905-0156596.
PR 04-OCT-1999; 9905-0157117.
PR 05-OCT-1999; 9905-0157753.

```
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159283.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 23-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
```

Query Match 17.2%; Score 450; DB 21; Length 472;

Best Local Similarity 26.6%; Pred. No. 6.3e-35;

Matches 134; Conservative 90; Mismatches 217; Indels 62; Gaps 15;

```
OY 6 LLASALCFVPLLLFLALKLMWLYCVSSRDRSCALPPPGTMOFPFGETLQNV---- 61
DB 9 llsiaagflll|-----rttyrmglppsglpligetfqligayk 52

OY 62 -LQRRKFLQMRKRYGYLYKTHLFGPTVRVAGADNVRRIILGEHRLVSVMPASVRTIL 120
DB 53 tenpepfidervarysvfmthlfgpelfsadbetrfrvlnegklfecsyasicnll 112

OY 121 GAGCLSNLHDSHKOKRVIMQAFSRBALQCYLV-IAEYVSSCLEQWLSGGERGLVYP 179
DB 113 ghslllmlkgsllhkrmlslmsfanssllkdhlnldrlvrfnldsws---rvllme 168

OY 180 EYKRIIMFRIARILILGCEPGAGGEDQOLVEAFEEEMTRMLFSLPIDVPPSGLYR-GVK 238
DB 169 eakktitfelvtqklnstfp-----geweslrkeyllviegffslplpl-fsttyrkaig 222

OY 239 AKNLIHARIENIRAKIRRLQATEPDG--CKDALQLLI--EHSWGERGLDMQALKQSS 294
DB 223 arr-----kvaaellyvmkkrreeeegeaerkkdmlaalaaaddgfsdeei|df-----l 272

OY 295 YELLFGHEHTASATSLITLYGLVPHVLOKVRREIRKSGILCKSNQDNKLDMETLEQJK 354
DB 273 vallyagyeltslmtlavkllteplalaqkee--hekllramksdyslswsdykscmp 330

OY 355 YIGCVIKETLRLNPPVPGCFRVALKTFELNGVOIPKGNVIVYSICDTHDVAIDFTNKEEF 414
DB 331 ftgcgvnetllvaanligvfrtamtdeveikykllpkgykwkvsftravhldpnhfkdartf 390

OY 415 NPDRE----YVHPEDASRFSFIPGGGIRSCVGEKFAKILKLTFTVELARHCDWOLLNG 470
DB 391 npwrwgsnsvttgpnsv-----ftpfggprlcpayelaravalsvflhr|lvlgfsw----- 441

OY 471 PPTMKTSPTVYVDNLPAFTTF 493
DB 442 vpaegdklvfptlrrtkryplf 464
```

Search completed: November 6, 2001, 13:25:27
Job time: 12655 sec

OM of: US-09-668-482-4 to: EST:* out_format : pfs

Date: Nov 6, 2001 2:26 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frame_p2n.model -DEV=xlip
-Q=c9n2.1/USPTO_spool/US09668482/runat.05112001.133353.10873/app-query.fasta.1.1680
-DB=EST -OFTM=fastap -SUFFIX=est -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000 -GAPOP=4.500
-GAPPEXT=0.050 -XGAPOP=10.000 -XGAPPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELPEXT=7.000 -START=1 -MATRIX=blissum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
-MAXLEN=2000000000 -USER=US09668482.c9n1.1.5307 -NCPU=6
-ICPU=3 -LONLOG -NO_XLIPX -WAIT -THREADS=1

Search information block:

Query: US-09-668-482-4

Query Length: 497

Database sequences:

Database sequences: 10228115
Database Length: 431459454
Search time (sec): 2793.150000

score_list:

Sequence	Strd	Orig	Zscore	EScore	Len	Documentation
gb_est29:AL539668	+	1431.50	2868.54	1.2e-150	974	AL539668 AL539668 LTI_FL013_FBT
gb_est29:AL539445	+	1376.00	2757.18	1.9e-144	893	AL532445 AL532445 LTI_NFL001_NE
gb_est29:AL532444	-	1313.00	2627.94	2.9e-137	1028	AL532444 AL532444 LTI_NFL001_N
gb_est18:AM513600	+	1045.00	2091.25	2.3e-107	618	AM513600 x047b06.y1 NCI_CGAP_U
gb_est29:AL539667	-	1024.00	2045.01	8.7e-105	847	AL539667 AL539667 LTI_FL013_FBT
gb_est17:BE236243	+	881.00	1760.81	5.8e-89	537	BE236243 143359 MARC_TFIG_Sus
gb_est17:BE358034	+	852.00	1702.05	1.1e-85	539	BE358034 602045632F1 NCI_CGAP_U
gb_est18:BF353034	+	852.00	1701.51	1.2e-85	564	BF353034 602073982F1 NCI_CGAP_U
gb_est14:BF353769	+	840.00	1674.73	3.6e-84	696	BF353769 602073982F1 NCI_CGAP_U
gb_est14:BB189825	+	823.00	1644.84	1.7e-82	474	AA239785 mx80a03.y1 Soares_mous
gb_est14:AA339785	+	769.00	1533.84	2.6e-76	545	BF336872 602027980F1 NCI_CGAP_U
gb_est18:BF236872	+	763.00	1519.49	1.6e-75	646	BF236872 717812.x1 Soares_NSL
gb_est18:BF055367	-	751.50	1495.12	3.7e-74	720	AM174347 f142a06.y1 Sugano_Kaw
gb_est14:AM174347	+	731.50	1455.49	5.9e-72	669	AM174347 f142a06.y1 Sugano_Kaw
gb_est15:AM765767	+	691.50	1375.01	1.8e-67	640	AM765767 fc24h03.y1 Zebrafish_W
gb_est23:AI667038	+	680.50	1355.44	2.2e-66	509	AI667038 fc24h03.y1 Zebrafish_W
gb_est107:R51129	+	673.50	1339.91	1.6e-65	571	R51129 y971b05.y1 Soares_infant
gb_est107:BG515142	+	673.50	1336.14	2.6e-65	662	BG515142 dad60f10.y1 Wellcome_C
gb_est143:AM128823	+	642.50	1276.51	5.5e-62	597	AM128823 fe37g11.y1 Zebrafish_W
gb_est102:BG578446	+	642.00	1277.18	5.5e-62	522	BG578446 de99c02.y1 Wellcome_C
gb_est19:BG364937	+	624.00	1241.33	3.0e-60	496	BF047649 dc80h02.y1 NICHD_XGC_C
gb_est18:BF047649	+	617.00	1226.78	3.3e-59	512	BF047649 dc80h02.y1 NICHD_XGC_C
gb_est18:BF426259	+	612.00	1215.54	1.4e-58	563	BF426259 df69e06.y1 Xenopus_lag
gb_est18:AM595907	+	588.00	1153.32	1.1e-55	765	AM595907 da29g11.y2 Xenla_131d
gb_est143:AM1771465	-	582.00	1153.32	1.1e-55	765	AM1771465 f142a06.y1 Sugano_Kaw
gb_est14:AM244588	-	579.00	1149.38	6.7e-55	532	AM244588 BR_EMD09D4ps2_Bain_Ka
gb_est14:AI1004222	+	579.00	1149.38	6.7e-55	532	AI1004222 ou03e01.y1 Soares_tota
gb_est12:AI584636	+	563.00	1117.03	4.2e-53	530	AI584636 fb81e05.y1 Zebrafish_W
gb_est17:AM460650	-	556.00	1099.95	4.2e-52	731	AM460650 da29g11.y1 Xenla_131d
gb_est17:AM460650	-	552.00	1090.94	2.6e-48	687	AM460650 df69e06.y1 Xenopus_lag
gb_est17:BG161868	+	521.50	1036.08	1.4e-48	408	BG161868 df69e06.y1 Xenopus_lag
gb_est107:R21282	+	508.00	1002.87	9.7e-47	671	R21282 y947d12.x1 Soares_infant
gb_est15:AM767659	+	501.00	992.36	5.9e-46	492	AM767659 da77a02.x1 Harland_sta
gb_est14:AM668461	+	499.00	986.82	5.9e-46	471	AM668461 AV668461 Sugano_Kaw
gb_est14:AM668503	+	499.00	986.82	5.9e-46	471	AM668503 AV668503 Sugano_Kaw
gb_est17:BE188917	-	475.00	968.94	4.6e-43	623	BE188917 db61e05.y1 Wellcome_C
gb_est14:AM199734	+	472.50	927.27	1.6e-42	919	AM199734 da6604.y1 Xenopus_lag
gb_est143:AM128313	-	468.50	922.97	2.7e-42	666	AM128313 fe37g11.y1 Zebrafish_W
gb_est12:AI545454	-	466.00	915.98	6.7e-42	784	AI545454 fb81e05.y1 Zebrafish_W
gb_est14:AM640733	+	463.50	918.43	4.9e-42	415	AM640733 b198f01.y1 Wellcome_C
gb_est101:BG487335	+	442.00	870.88	2.2e-39	583	BG487335 dad5d03.y1 Wellcome_C
gb_est15:AM766054	+	424.00	833.90	2.5e-37	664	AM766054 da81f06.y1 Harland_sta
gb_est101:BG487338	-	421.00	828.25	5.1e-37	588	BG487338 dad5d04.x1 Wellcome_C

gb_est15:BE506442 + 407.50 805.32 9.7e-36 405 ! BE506442 db83b11.y1 Wellcome
gb_est101:BG486108 + 404.00 793.73 4.3e-35 593 ! BG486108 dad60f10.x1 Wellcome
gb_gss5:CNS05P00 + 402.50 783.90 1.5e-34 1054 ! AL348447 Tetraodon nigrovir
gb_est19:BG348623 + 395.50 782.18 1.9e-34 367 ! BG348623 dad18c02.y1 Wellcome
seq_name: gb_est29:AL539668

seq_documentation block:

LOCUS AL539668 974 bp mRNA EST 16-FEB-2001
DEFINITION AL539668 LTI_FL013_Fbrn1 Homo sapiens cDNA clone CS0DF036YG04 5
prime, mRNA sequence.
ACCESSION AL539668
VERSION AL539668.1 GI:12869097
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 974)
L1.W.B., Gruber,C., Jesse,J. and Polayes,D.
Full-length cDNA libraries and normalization
Unpublished (2001)
CONTACT: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr. Web : www.genoscope.cns.fr.

FEATURES

source
1..974
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="LTI_FL013_Fbrn1"
/dev_stage="pooled tissue from post conception fetuses (20
week, 24 week and 26 week)"
/tab_host="DH10B"
/note="Organ: Fetal brain; Vector: PCMVSPORT 6; 1st strand
cDNA was primed with a Notti-oligo(dT) primer. First strand
end enriched, double-stranded cDNA was digested with Not I
and cloned into the Not I and Eco RV sites of the
PCMVSPORT 6 vector. Library was constructed by Life
Technologies. Contact : Feng Liang Life Technologies, a
division of Invitrogen 9800 Medical Center Drive Rockville
Maryland 20850, USA Fax : (1) 301 610 8371 Email :
filiang@life.com URL :
http://fulllength.invitrogen.com"

BASE COUNT 223 a 269 c 305 g 177 t
ORIGIN

alignment_scores:
Quality: 1431.50 Length: 300
Ratio: 4.853 Gaps: 1
Percent Similarity: 98.333 Percent Identity: 96.667

alignment block:

US-09-668-482-4 x AL539668 ..

Align seg 1/1 to: AL539668 from: 1 to: 974

54 Phegylglturthleuglmetvalleuglnargargylspheleuglme 70
||||| ||| : : : : : ||||| ||||| ||||| ||||| |||||
77 TTGGGCGACT...ATTGCTGACCAAGCGAGGAGATTCGCGCAT 123
70 tlyatgargylstyrgylpheylytyrlystfhnlisleupheglyarp 87
||||| ||| : : : : : ||||| ||||| ||||| ||||| |||||
124 GAAGCGCGAATAATGCGCTCATCTCAAGACCATCTGTCGCGCGC 173
87 rothrvatargvalmetgylvalaspasvalargargyleleugly 103
||||| ||| : : : : : ||||| ||||| ||||| ||||| |||||
174 CCACCGTCGAGGATGCGCGCGGACCAATGCGCGCATCTGCTGGA 223
104 Aspaspargvalservalistripvalaservalargthrllele 120
||||| ||| : : : : : ||||| ||||| ||||| ||||| |||||


```

380 TCCGAGCAGCAGCTTGTGAGAGCCCTTCGAGAAATGACCCGCATCTCTT 429
222 eSerLeuProIleAspValProPheSerGlyLeuTyrArgGlyMetLysA 239
430 CTCCTCCCAATCGACGTCGCCCTTCACGGGGCTGTACCGGGCATGAAG 479
239 laaGAsnLeuIleHisAlaArgIleGluGlnAsnIleArgAlaLysIle 255
480 CGCGAACCCTATTCACGCGCGCATCGACGAGAAACATTCGCGCAACATC 529
256 CysGlyLeuArgAlaSerGluAlaGlyGlnGlyCysLysAspAlaLeuG 272
530 TGGCGGCTCGCGGCATCCGAGCGCGGCGCTGCAAAAGACGGCTGCA 579
272 nLeuLeuIleGlnHisSerTyrPgluArgGlyGluArgLysAspMetGlnA 289
580 GCTGTGTGTCAGACATCGTGGGAGAGGGGAGCGGCTGGACATGACG 629
289 laLeuLysGlnSerSerThrGluLeuLeu..PheGlyGlyHisGluThrTh 305
630 CACTTAAGCAATCTTCAACCACTCCTCTTTGGAGSACGAGAAACAC 679
305 rAlaSerAlaAlaThrSerLeuIleThrTyrLeuGlyLeuTyrProHisV 322
680 GGCCAGTCGACCAACATCTGTATCATTACCTGGGGCTTACCCACATG 729
322 alLeuGlnLysValArgGluGluLeuLysSerLysGlyLeuLeuCysLys 338
730 TTCCTCAAAAGTGCAGAAAGAGCTGAAGATGAAGGTTACTTTGACAG 779
339 SerAsnGlnAspAsnLysLeuAspMetGlnIleLeuGlnLeuLysTyr 355
780 AGCAATTCAGACAAACACTTGCACATGGAATTTTGGACACACTTAATA 828
355 rIleGlyCysValIleLysGluThrLeuArgLysAsnProProValProG 372
829 CATCGGCTGTGTATTAGAGACCTTCGACTGATCCCACT...TCAG 875
372 lyGlyPheArgValAla 377
876 GAGGGTTTCGGGTTGCT 892

```

seq_name: gb_est29:AL532444

seq_documentation_block:

LOCUS AL532444 1028 bp mRNA EST 13-FEB-2001

DEFINITION AL532444 L1_NFL001_NBC4 Homo sapiens cDNA clone CS0DM014YL13 3

ACCESSION AL532444

VERSION AL532444.1 GI:12795937

KEYWORDS

SOURCE

ORGANISM

human.

Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1. (bases 1 to 1028) Li, W.-B., Gruber, C., Jessee, J. and Polayes, D.

Full-length cDNA libraries and normalization

Unpublished (2001)

JOURNAL

COMMENT

Genoscope - Centre National de Sequencage

BP 191 91006 EVRY cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES

source

1. 1028

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="CS0DM014YL13"

/clone_11b="L1_NFL001_NBC4"

/sex="male"

/tissue_type="neuroblastoma cells"

/lab_host="DH108"

/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA

was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Peng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : <http://fulllength.invitrogen.com>

BASE COUNT 265 a 244 c 225 g 282 t 12 others

alignment_scores:

Quality: 1313.00 Length: 282

Ratio: 4.95 Gaps: 3

Percent Similarity: 93.972 Percent Identity: 91.489

alignment_block:

US-09-668-482-4 x AL532444/rev ..

Align seg 1/1 to reverse of: AL532444 from: 1 to: 1028

```

217 MetThrArgAsnLeuPheSerLeuProIleAspValProPheSerGlyLe 233
||||| :|: ||||| ||| |||:|
1005 ATGACCCGA...ATCTCTCTCGCTCCCATCGACTGCCCT...TCAGCGGC 962

```

```

233 uTyrArgGlyMetLysAlaArgAsnLeuIleHisAlaArgIleGluGlnA 250
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
961 TGTACCGGCGCATGTAGCGGAACCTTAATTAACGCGCGATCGACCA 912

```

```

250 snIleArgAlaLysIleCysGlyLeuArgAlaSerGluAlaGlyGlnGly 266
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
911 AATATCGCGCCAGATGTGCGGCTGCGGCATCCGAGCGGCGCAGGGC 862

```

```

267 Cys.LysAspAlaLeuGlnLeuLeuIleGluHisSerTyrPgluArgGly 283
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
861 TGCAAAAGACGCGCGTGCAGCTGTGATCGAGCACGTGGAGAGGGAG 812

```

```

283 LuArgLeuAspMetGlnAlaLeuLysGlnSerThrGluLeuPhe 299
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
811 AGCGGTGAGACATGCAGCACTTAAGCAATCTTAACCGCAACTCTCTTT 762

```

```

300 GlyLysGlnGluThrThrAlaSerAlaAlaThrSerLeuIleThrTyrI 316
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
761 GGAGGACAGAAACACAGCGCAGGACGACCATCTCTGATMACTTACCT 712

```

```

316 uGlyLeuTyrProHisValLeuGlnLysValArgGluGluLeuLysSerL 333
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
711 GGGGCTCTACCCACATGTCTCCGAAAGTGCAGAAAGACTGAAGATA 662

```

```

333 ysglyLeuLeuCysLysSerAsnGlnAspAsnLysLeuAspMetGluIle 349
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
661 AGGTTTACTTTTGCAGAACATCAAGCAACAAGTTGGACATGGAAT 612

```

```

350 LeuGluGlnLeuLysTyrIleGlyCysValIleLysGluThrLeuArgI 366
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
611 TTGGAACAACTTAATACATCGGCTGTGTTTAAAGAGACCTTGCACT 562

```

```

366 uAsnProProValProGlyLysPheArgValAlaLeuLysThrPheGluL 383
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
561 GAATCCGCCAGATTCCAGAGAGGTTGGGTTGCTGCAACATTTTGAAT 512

```

```

383 euAsnGlyTyrGlnIleProLysGlyTyrPasnValIleTyrSerIleCys 399
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
511 TAAATGATACCATTCACATTCAGAGGCTGGAATGTATCTACAGATCTGT 462

```

```

400 AspThrHisAspValAlaGluIlePheThrAsnLysGluLupheAsnPr 416
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
461 GATACCTCATGTGTGGCAGAGATCTTCAACACAGAGAAATTTAATCC 412

```

```

416 oAspArgPheSerAlaProHisProGluAspAlaSerArgPheSerPhe 433
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

411 TGACCGATTCATGCTGCCTCACCCAGAGATGCATCCAGTTTCAGCTCA 362
 433 leprohegilyglyleuargsercysvalglysghlphalalays 449
 361 TTCATTGGAGAGCCTTAGAGCTGTAGCAAGAAATTTGCCAAA 312
 450 lileuleuysilphehthvalgluleualarhishscysasptprgl 466
 311 ATTCTCTCAAAATATTACAGTGTAGAGTGTSCAGGCACTTTSACTGCM 262
 466 nleuleuansglyproprohmetylnhserprothrvaltyrprov 483
 261 GCTTCTAAATGGACTCTACATGAAGAAWCAAGCCACCGTATCTCTG 212
 483 alaspaanleuproalarphethrhishphehsglyluile 497
 211 TGGCATCTCCCTGCAGATTCACCATTTCCATGGGGAATC 168
 seq_name: gb_est48:AM513600

seq_documentation_block:

LOCUS AM513600 618 bp mRNA EST 03-MAR-2000
 DEFINITION x047b06.x1 NCI_CGAP_Ut1 Homo sapiens cDNA clone IMAGE:2707091 3'
 similar to SW:CP26_HUMAN 043174 CTTCCHROME P450 26 ;, mRNA
 sequence.

ACCESSION AM513600 GI:7151678
 VERSION AM513600
 KEYWORDS EST.

SOURCE human.

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE

1 (bases 1 to 618)

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index

Unpublished (1997)

JOURNAL

CONTACT: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.

Emmert-Buck, M.D., Ph.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

image.llnl.gov/image/html/resources.shtml

Possible reversed clone: similarity on wrong strand

Seq primer: -400P from gibco

High quality sequence stop: 440.

Location/Qualifiers

1..618

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:2707091"

/clone_lib="NCI_CGAP_Ut1"

/tissue_type="well-differentiated endometrial

adenocarcinoma, 7 pooled tumors"

/lab_host="DH10B"

/note="Organ: uterus; Vector: pCMV-SPORT6; site_1: SalI;
 site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
 Average insert size 1.75 kb. Life Technologies catalog #:
 11538-014"

BASE COUNT 111 a 191 c 199 g 115 t 2 others

ORIGIN

alignment_scores:
 Quality: 1045.00 Length: 206
 Ratio: 5.148 Gaps: 0
 Percent Similarity: 98.544 Percent Identity: 98.058

alignment_block:
 US-09-668-482-4 x AM513600 ..

Align seg 1/1 to: AM513600 from: 1 to: 618

34 SerGlyArgAspArgSerCysAlaLeuProLeuPropGlyThrMetG1 50
 1 AGCGGCCGACACCGAGTTGTGCCCTCCATTGCCCGGGGACTATGG 50
 50 yPheProPhePheGlyGlnThrLeuGlnMetValLeuGlnArgArgLysP 67
 51 CTTCCTCCCTTTCTTGGGGAACCTTGACATGTAAGTCAAGAGGAGT 100
 67 heuGlnMetLysArgArgLysTyrGlyPheIleTyrLysThrHisLeu 83
 101 TCCTGCAATGAGGCGAGAAATACGCTTCACTTCAACAAAGCATCTG 150
 84 PheGlyArgProThrValArgValMetGlyAlaAspAsnValArgArg1 100
 151 TTGGGGCGGCCCGACGATACGGGTGTGTGGCGGCAATGTGGCGCAT 200
 100 eLeuLeuGlyAspAspArgLeuValSerValHisTrpProAlaSerVal 117
 201 CTTCCTCGAGAGACACCGCGCTGTGTGTCTCCTGCTGACGAGCTCGTGC 250
 117 rGThrIleLeuGlySerGlyCysLeuSerAsnLeuHisAspSerSerHis 133
 251 GCACCATTTCTGGATCTGTGCTGCTCTTACCTGACGAGCTCCTCGAC 300
 134 LysGlnArgLysLysValIleMetArgAlaPheSerArgIuaIaLeuG1 150
 301 AAGCAGGCGCAAGAGGTATATGCGGGCCCTTACGCGGAGGCACTGCA 350
 150 uCysTyrValProValIleThrGlnGluValGlySerSerLeuGlnT 167
 351 ATGCTACGTGCGGTGATCACCGAGAAAGTGCGCAGAGCTGAGAGAGT 400
 167 rPheSerCysGlyGluArgLysLeuValTyrProGluValLysArg 183
 401 GCGTACGCTGCGGCGAGCGCGCTCCTGCTGCTACCCGAGGTAAAGGC 450
 184 LeuMetPheArgIleAlaMetArgIleLeuLeuGlyCysGluProGlnLe 200
 451 CTCATGTTCCGAATGCCATGCGCATCTACTGTGTCGGAACCCAACT 500
 200 uAlaGlyAspGlyAspSerGlnGlnGlnLeuValGluAlaPheGlnGlu 217
 501 GCGGGCGAGCGGAGCTCCGAGCAGAGCTTGTGAGAGCCCTTCAGAGAAA 550
 217 eThrArgAsnLeuPheSerLeuProIleAspValProPheSerGlyLeu 233
 551 TGACCGCATCTCTTCTGCTGCTCCATTCAGCGTGCCTTCAGCGGCTG 600
 234 TyrArgGlyMetLysAla 239
 601 TACCGGGCGCATGAAGGCN 618
 seq_name: gb_est29:AL539667

seq_documentation_block:
 LOCUS AL539667 847 bp mRNA EST 16-FEB-2001
 DEFINITION AL539667 UT1.FL013.Fbrn1 Homo sapiens cDNA clone CS0DF036YG04 3
 prime, mRNA sequence.
 ACCESSION AL539667 GI:12869095
 VERSION AL539667.1
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE 1 (bases 1 to 847)
 AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
 TITLE Full-length cDNA libraries and normalization

JOURNAL Unpublished (2001)

CONTACT: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: segreff@genoscope.cns.fr, Web : www.genoscope.cns.fr.
Location/Qualifiers

FEATURES

source
1. 847
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="LTI_FLO13_FBrn1"
/dev_stage="pooled tissue from post conception fetuses (20 week, 24 week and 26 week)"
/lab_host="DH10B"
/note="Organ: Fetal brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville , Maryland 20850, USA Fax : (1) 301 610 8371 Email : liang@lifetech.com URL : <http://fulllength.invitrogen.com>"

BASE COUNT 197 a 190 c 204 g 229 t 27 others
ORIGIN

alignment_scores:

Quality: 1024.00 Length: 231
Ratio: 4.995 Gaps: 0
Percent Similarity: 88.745 Percent Identity: 85.281

alignment_block:

US-09-668-482-4 x AL539667/rev ..

Align seg 1/1 to reverse of: AL539667 from: 1 to: 847

```

266 GTCGTCCTCAAGACGCGCTGCTGATCGACCTGCTGAGAGGGG 797
|||||
282 TGTATGTCCTCAAGACGCGCTGCTGATCGACCTGCTGAGAGGGG 797
|||||
796 AGAGGGGCTGAGCATGCGACGCTAAGCATCTTCAACGACCTCTCT 747
|||||
299 heglylghisgluThrAlaSerAlaIarThrSerleuIleThrTyr 315
|||||
746 TTGGAGGACAGCAACGAGGCGACGACGACATCTGTGATCCTTAC 697
|||||
316 leuglyleuTyrrProHisValleuGlnIysValaIargIuIleuLys 332
|||||
696 CTGGGGGCTCCACCCACGCTCCCGAAGAGCGAGAGAGCTGAAAG 647
|||||
332 rlysglyleuLeuCylysserAsnGlnAspAsnIysLeuAspMetGlu 349
|||||
646 TTAGGCTTACTTCTTCAAGMGCAATCAAGACACAAAGTTGACATG 597
|||||
349 leleugluGlnleuLysTyrrIleGlyCysValIleIysgluThrLeu 365
|||||
596 TTTTGGAACTCAATTCATTCGCGGTGTATTAAGAGAGACCCCTCG 547
|||||
366 leuAsnProProValProGlyIlePheArgValAlaIleuLysThr 382
|||||
546 CTGAACCCCGCCGCGAGGAGGTTTCGCGGNNCTCCGAASMCCTT 497
|||||
382 uleuAsnGlyTyrrGlnIleProIysGlyTrpAsnValIleTyrrSer 399
|||||
496 ATCAATATGATCCGAGCAGCCAGAGGCTGGAATGTATCTACATAT 447
|||||
399 yAspThrHisAspValAlaIlePheThrAsnIysgluIleuPheAsn 415
|||||
446 GTGATACCATGATGTGGCAGAGATCTTCAACACAGAGAAATTTAAC 397

```

```

416 ProAspArgPheSerAlaProHisProGluAspAlaSerArgPheSer 432
|||||
396 CCGACCGATTCATGCTGCTCACCAGAGAGACGCCCGGCTCAGCCC 347
|||||
432 eilLeuProPheGlyGlyIleuArgSerCysValIleGlyGluPheAla 449
|||||
346 CATCCCATTTGGAGGAGGCGCTTAGAGCTGTGTAGCAAAAGATTCG 297
|||||
449 ysilLeuLeuLysIlePheThrValGluLeuAlaArgHisCysAsp 465
|||||
296 AAATCCCGCCCAAAATTTACAGTGTGAGCTGSCAGGCAATGTGACT 247
|||||
466 GlnLeuLeuAsnGlyProProThrMetLysThrSerProThrValTy 482
|||||
246 CCSCBCTTAATGAGACCTCTACAGAAACCAACGCCGCTGATTC 197
|||||
482 oValAspAsnLeuProAlaArgPheThrHisPheHisGlyGlu 496
|||||
196 TGTGGACATCTCCCTGCAAGATTCACCATTCCTCATGGGGA 154
|||||
seq_name: gb_est72:BE236243

```

seq_documentation_block:

LOCUS BE236243 537 bp mRNA EST 10-JUL-2000
DEFINITION 143959 MARC 1Pig Sus scrofa cDNA 5', mRNA sequence.
ACCESSION BE236243
VERSION BE236243.1 GI:9020961
KEYWORDS EST.
SOURCE pig.
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

REFERENCE

AUTHORS

TITLE

Design and use of two pooled tissue normalized cDNA libraries for EST discovery in swine
Unpublished (2000)
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred
v0.980904.e. Vector identified by cross_match with the minscore 18
and -mismatch 12 options.
PCR Primers
FORWARD: AGGAACAGCTTATGACAT
BACKWARD: GTTTCCAGTCAGCAGC
Plate: 85 row: B column: 21
Seq primer: ATTAGTGACACTATAG.
Location/Qualifiers

JOURNAL

COMMENT

FEATURES

source

1. 537
/organism="Sus scrofa"
/db_xref="taxon:9823"
/clone_lib="MARC 1Pig"
/issue_type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV SPORT6; Site 1: XbaI; Site 2: XhoI;
library made from pooled tissue from day 11, 13, 15, 20,
and 30 embryos."
BASE COUNT 96 a 163 c 174 g 104 t
ORIGIN

alignment_scores:

Quality: 881.00 Length: 177
Ratio: 5.065 Gaps: 0
Percent Similarity: 98.305 Percent Identity: 93.785

alignment_block:
US-09-668-482-4 x BE236243 ..

Align seg 1/1 to: BE236243 from: 1 to: 537

```

25 11elysleuTPaspLeuTyrcysValserGlyarGasparsercysAl 41
|||||
1 ATCAAACTCTGGGACCTGTACTGCTGAGCAGCCGGAGCCGAGCTGCAC 50
41 aleuProleuProProGlyThrMetGlyPheProPhehegIyIunrL 58
|||||
51 CCTTCCTTGGCCCCCTGGAACTATGGCTCCCTCTTCTTGGGGAGACAT 100
58 euglMetValleuGlnArqArqLysPheleuGlnMetLysArqArqLys 74
|||||
101 TGCAGATGCTGCTACAGCGAAGAGATTCTGACATGACGCGCAGAA 150
75 TyrglyPheileTyrcysThrHisLeuPheglArqProThrValArqVa 91
|||||
151 TACGTTTCATCTACAAAGACCATCTGTCGGGAGGCCACGGTCGGGT 200
91 lMetGlyAlaspasValArqArqIleleuLeuGlyaspasparleu 108
|||||
201 GATGGGTGACAGACACGTCGGCGCATCTTCTCGGGAGAACACCGCTCG 250
108 alserValHisTrpProAlaserValArqThrIleleuGlyserGlyCys 124
|||||
251 TGTGCTTCACATGGCGCGCGTCGTGCGCCACGATCTGGCTGTGGCTGC 300
125 leuSerAsnLeuHisAspSerSerHisLysGlnArqLysValIleMe 141
|||||
301 CTCTCAACCTGCACGACTCTCCGACAAAGCAGCGCAAGAGTATAT 350
141 tArqAlPheSerArqGlnAlaleuGlnCysTyrcysValProValIleThng 158
|||||
351 GCAGGCTTCAGCCGCGAGCGCTCCAGTCTGCTGCTGCGGTGATCGAG 400
158 lugiValGlySerSerleuGlnTrpleuSerCysGlyIunrGly 174
|||||
401 AGAAGTGGACAGTTCGCTGGAGCAGTGGCTGGGAGAGCGCGC 450
175 leuLeuValTyrcysProGlnValLysArqLeuMetPhearqIleAlaMetAr 191
|||||
451 CTCTGTGTTTACCCCGAGTGTAAGCGCTCATGTTCGCGATCGCATCG 500
191 gIleLeuLeuGlyCysGlnProGlnLeuAla 201
|||||
501 CATCTGCTGGGCTGCCAGCCCGGCTGGCG 531

```

seq_name: qb_est87:BF385034

seq_documentation_block: 539 bp mRNA EST 27-NOV-2000
 LOCUS BF385034 602045632rf1 NCI_CGAP_L19 Mus musculus cDNA IMAGE:4195422 5',
 DEFINITION mRNA sequence.
 ACCESSION BF385034
 VERSION BF385034.1 GI:11366326
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus.

REFERENCE
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
 TITLE 1 (bases 1 to 539)
 JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
 COMMENT Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cga@db-rrmail.nih.gov
 Tissue Procurement: Jeffrey E. Green, M.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LHAM9530 row: 0 column: 07
 High quality sequence stop: 539.
 Location/Qualifiers
 1..539
 /organism="Mus musculus"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone="IMAGE:4195422"
 /clone_id="NCI_CGAP_L19"
 /lab_host="DH10B (T1 phage-resistant)"
 /note="Organ: liver; Vector: pCMV-SPOrt6; Site:1: NotI;
 Site:2: SalI; Cloned unidirectionally. Primer: Oligo dt.
 Average insert size 1.9 kb. Constructed by Life
 Technologies. Note: this is a NCI_CGAP library."

BASE COUNT 86 a 160 c 181 g 112 t
 ORIGIN

alignment_scores:
 Quality: 852.00 Length: 171
 Ratio: 5.041 Gaps: 0
 Percent Similarity: 98.830 Percent Identity: 94.152

alignment_block:
 US-09-668-482-4 x BF385034 ..

Align seg 1/1 to: BF385034 from: 1 to: 539

```

1 MetGlyLeuProAlaLeuLeuAlaserAlaleuCysThrPheValleuPr 17
|||||
25 ATGGGGCTCCCGCGCGCTGCGCCAGTGGCTGCTGCTGCTGCTGCTGCTG 74
17 OleuLeuLeuPheLeuAlaAlaIleLysLeuTrpAspleuTyrcysVal 34
|||||
75 GCTGCTGCTCTTCCGCGCGGCTCAAGCTCTGGAGCCTGTACTGTGTA 124
34 erglyArqAspArqSerCysAlaleuProleuProProGlyThrMetLys 50
|||||
125 GCAGCGCGATCGCAGCTGCGCCCTCCCTTGCCTCCCGGTAACCTAGGCG 174
51 PheProPhePheGlyGlnThrLeuGlnMetValleuGlnArqArqLysPhe 67
|||||
175 TTCCCATCTTTGGGGAACAATGACAGTGGCTGCTGACGCGAGAAATT 224
67 eleuGlnMetLysArqArqLysTyrcysPheIleTyrcysThrHisLeuP 84
|||||
225 TGTGAGATGAAGCGCAGAAATAGCGCTCATCTACAAAGACGCAATCTGT 274
84 heglYArqProThrValArqValMetGlyAlaspasValArqArqIle 100
|||||
275 TTGGGCGGCCACAGGTCGGGTGATGGCGCGGATATGTGGGGGCAATC 324
101 leuLeuGlyAspAspArqLeuValserValHisTrpProAlaserValAr 117
|||||
325 TTGCTGGAGAGACACCGGTGTGTGCGGTGACACGCGCGCGGTGCG 374
117 gThrIleleuGlySerGlyCysLeuSerAsnLeuHisAspSerHisL 134
|||||
375 CACCATCTGGGCGCTGCTGCTCCACACCTGCACATTCCTCGGCA 424
134 ysglArqLysLysValIleMetArqAlaPheSerArqGlnAlaleuGln 150
|||||
425 AGCAGCGAAGAAAGTGTATTCAGAGCCTTCAGCGCGGAGGACCTCAG 474
151 CysTyrcysValProValIleThrGlnGlnValGlySerSerleuGlnTrp 167
|||||
475 TGTACAGGCGCGGTGATCGCTGAGAGAGTACGAGTGTCTGAGGAGATG 524
167 pleuSerCysGly 171
|||||
525 GCTAAGCTGCGGC 537

```

```

seq_name: gb_est89:BF533769
seq_documentation_block:
LOCUS      BF533769          564 bp      mRNA      EST      11-DEC-2000
DEFINITION 602073982F1 NCI_CGAP_L19 Mus musculus cDNA clone IMAGE:4210893 5',
            mRNA sequence.
ACCESSION  BF533769
VERSION    BF533769.1  GI:11621132
KEYWORDS   house mouse.
SOURCE     Mus musculus.
ORGANISM   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 564)
AUTHORS   NIH-MGC http://mgi.nci.nih.gov/.
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LIML at:
            http://image.llnl.gov
            Plate: L1A99779 row: c column: 22
            High quality sequence stop: 564.
FEATURES
    source
        location/Qualifiers
            1..564
            /organism="Mus musculus"
            /strain="FVB/N"
            /db_xref="taxon:10090"
            /clone_1ib="IMAGE:4210893"
            /lab_host="NCI_CGAP_L19"
            /note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI;
            Site_2: SalI; Cloned unidirectionally. Primer: oligo dt.
            Average insert size 1.9 kb. Constructed by Life
            Technologies. Note: this is a NCI_CGAP library."
BASE COUNT      89 a      172 c      189 g      114 t
ORIGIN
alignment_scores:
    Quality: 852.00      Length: 171
    Ratio: 5.041      Gaps: 0
Percent Similarity: 98.830      Percent Identity: 94.152
alignment_block:
US-09-668-482-4 x BF533769 ..
Align seg 1/1 to: BF533769 from: 1 to: 564
1 MetGlyLeuProAlaLeuLeuAlaSerAlaLeuCysThrPheValLeuPr 17
|||||
50 ATGGGGCTCCGGCGCTGCTGGCCAGTGGCTTGCACCTTGTCTGCTGCC 99
|||||
17 oleuLeuLeuPheLeuAlaAlaAlaLeuLeuTrpAspLeuTrpCysValS 34
|||||
100 GCGGCGCTCTTCCTGGCGGCGCTCAAGCTGTGGACGTGACTGTGTCGA 149
|||||
34 erGlyArgAspArgSerCysAlaLeuProLeuProProGlyThrMetGly 50
|||||
150 GCAGCGCGCATGCGACGTCGCCCTCCCTCCGCCCGCGTACCAAGGGGC 199
|||||
51 PheProPhePheGlyGlyThrLeuGlnMetValLeuGlnArgArgLysPh 67
|||||
200 TTCCCATTTCTTGGGAAACATTGCGAGATGTGCTTCAACGGAGGAAGT 249
|||||
67 eLeuGlnMetLysArgArgLysTrpGlyPheIleTrpLysThrHisLeuP 84
|||||

```

```

250 TCTGCAGATGAAGCGGAGAAATACGGCTTCACTACAGACCACTCTGT 299
84 heGlyArgProThrValArgValMetGlyAlaAspAsnValArgArgIle 100
|||||
300 TTGGGGCGGCCACGCGTGGATGGCGCGATATGATGGCGGCATC 349
|||||
101 LeuLeuGlyAspAspArgLeuValSerValHisTrpProAlaSerValAr 117
|||||
350 TTCTGTGGAGAGACCGGTTGTGTGCTGCTGCTGCTGCTGCTGCTGCTG 399
|||||
117 gThrIleLeuGlySerGlyCysLeuSerAsnLeuHisAspSerSerHisL 134
|||||
400 CACCATCTTGGCGCGTGGCTGCTGCTTCCAACTGACAGATTCCTCGCACA 449
|||||
134 ysgIArgLysLysValIleMetArgAlaPheSerArgGluAlaLeuGlu 150
|||||
450 AGCAGCGAAAGAGGTATTATGACAGCCTTCAGCCGCGAGGACATCCAG 499
|||||
151 CysTrpValProValIleThrGluGluValGlySerSerLeuGluInTr 167
|||||
500 TGTCTAGTGGCCCGGTGATCGCTGAGAGAAATGACAGATGTTGTGTGAGCAGTg 549
|||||
167 PleuSerCysGly 171
|||||
550 GCTAAGCTCGCGGC 562
|||||
seq_name: gb_est71:BE189825
seq_documentation_block:
LOCUS      BE189825          696 bp      mRNA      EST      07-AUG-2000
DEFINITION db61c05.y1 Wellcome CRC psk egg Xenopus laevis cDNA clone
            IMAGE:3377480 5' similar to gb:gb1AF057566.1AF057566 Xenopus
            laevis retinolic acid converting enzyme (XENOPUS);, mRNA sequence.
ACCESSION  BE189825
VERSION    BE189825.2  GI:9729548
KEYWORDS   EST.
SOURCE     African clawed frog.
ORGANISM   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipridae;
            Xenopus laevis.
REFERENCE  1 (bases 1 to 696)
AUTHORS   Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L., Pape,D.,
            Martin,J., Wylie,T., Underwood,R., Theising,B., Bowers,Y., Person
            ,B., Gibbons,M., Harvey,N., Ritzer,E., Jackson,Y., McCann,R.,
            Waterston,R. and Wilson,R.
            WashU Xenopus EST project, 1999
            Unpublished (1999)
            On Jun 22, 2000 this sequence version replaced gi:8668718.
            Other_ESTs: db61c05.x1
            Contact: Sandy Clifton, Ph.D.
            WashU Xenopus EST project, 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Library constructed by N. Garrett, P. Lemaire, A.M. Zorn, and J.B.
            Gurdon (Wellcome/CRC Institute). DNA sequencing by: Washington
            University Genome Sequencing Center
            Clone distribution: Xenopus clones from this library are available
            through the I.M.A.G.E. Consortium/LIML at: info@image.llnl.gov
            Seq primer: -40RP from Gibco
            High quality sequence stop: 483.
FEATURES
    source
        location/Qualifiers
            1..696
            /organism="Xenopus laevis"
            /db_xref="taxon:8355"
            /clone_1ib="Wellcome CRC psk egg"
            /tissue_type="egg"
            /lab_host="DH10B (phage-resistant)"
            /note="Vector: pBluescript SK-; Site_1: NotI; Site_2:

```

ECORI; cDNAs were oligo-dT primed and directionally cloned. Library was constructed by N. Garrett, P. Lemaire, A.M. Zorn, and J.B. Gurdon (Wellcome/CRC Institute).
 Note: This is a Xenopus Gene Collection (XGC) library."

BASE COUNT 156 a 200 c 202 g 138 t
 ORIGIN

alignment_scores:
 Quality: 840.00 Length: 235
 Ratio: 4.118 Gaps: 4
 Percent Similarity: 86.809 Percent Identity: 71.064

alignment_block:
 US-09-668-482-4 x BE189825 ..

Align seg 1/1 to: BE189825 from: 1 to: 696

```

43  ProlenProProgllyThrMetGlyPheProPhePheGlyGluThrIleuG1 59
|||||
8  CCACTGCCCCCGGCGACTATGGGCGCTCTTCCGGGAGACTCTGCA 57
|||||
59  nmeValLeuGlnArgArgGlyPheLeuGlnMetLysArgArgLysTyrg 76
|||||
58  AATGTCCTCGAGAGGCGCAAGTCTCCAAATGAGCGTAGAAAGTACG 107
|||||
76  lYpHeileTyrlYsThrlsleuPheGlyArgpProThrlValArgValMet 92
|||||
108  GTCGACATCTACAGACCATCTGTTCGTTAGCCCGACGGTGGGTACG 157
|||||
93  GlyAlaAspAsnValArgArgIleLeuLeuGlyAspAspArgLeuValse 109
|||||
158  GGGGACAGAGACGTTCCGACATCTATTGGGGAGACCAACAGCTGTGTC 207
|||||
109  rValHlstrProAlaSerValArgThrIleLeuGlySerGlyCysLeus 126
|||||
208  GGGGACTGCGCGCCCTGCGGCGCAGATCCCAAGGGCGGCTGTGT 257
|||||
126  eArAsnLeuHlAspSerSerHlslsGlnArgLysLysValIleMetArg 142
|||||
258  CCACCGTGCAGACTCTGACACAGTACACCAAGAAAGTATTCACAA 307
|||||
143  AlaphSerArgLualAlaLeuGluCysTyrlValProValIleThrGluG1 159
|||||
308  GCGTTCGCCGAGATGCGCTCGACAAATTACGTGCCGACATGGAAGA 357
|||||
159  uValGlySerleuGlnGlnThrleuSerCysGlyLuarGlyLeul 176
|||||
358  GGTGAGACGCTGTGTAAACCTGTGGCTGACAGCGGCCCC...TGCCTGC 404
|||||
176  euValTyrlProGluValLysArgLeuMetPheArgIleAlaMetArgIle 192
|||||
405  TGTATGATCCCGCCATCAAGCGACTGATGTTCGCATATGCGCATGAAGCTC 454
|||||
193  leuLeuGlyCysGluProGlnLeuAlaGlyAspGlyAspSerGluGlnG1 209
|||||
455  CTGCTCGCTGCGATCCCGACGCGATG...GACAGCCCAACATGAGGAGAC 501
|||||
209  nLeuValGluAlaPheGluGluMetThrArgAsnLeuPheSerleuProI 226
|||||
502  GGTGCTCGAAGCCTTCGAGGAAATGACCCGAAATCTCTCTGTGCTGA 551
|||||
226  lAspValProPheSerGlyLeuTyrlArgGlyMetLysAlaArgAsnLeu 242
|||||
552  TTGATGTGCCATGTAGCGGCTCTACCGGGGCTCTGGGGCTAGAGACTT 601
|||||
243  lHlHlAlaArgIleGluGlnAsnIleArgAlaLysIleCysGlyLeuAr 259
|||||
602  ATTCAATGCGCGAATTGATGAAAAACATCGAAGAGAAAGCTG...CTAAG 645
|||||
259  gAlaSerGluAlaGlyGlnGlyCysLysAspAlaLeuGlnLeuLeuIleG 276
|||||
646  G.....GAACAGACGATATTGCGGGGATGCCCTCGACGCTGCTGATTG 689

```

276 luhis 277
 :
 690 ACTAT 694

seq_name: gb_est4:AA239785

seq_documentation_block:

LOCUS AA239785 474 bp mRNA EST 03-MAR-1997
 DEFINITION mx80a03.r1 Soares mouse NML Mus musculus cDNA clone IMAGE:692620 5' similar to TR:G1001252 G1001252 HYPOTHEICAL 50.6 KD PROTEIN ; mRNA sequence.

ACCESSION AA239785
 VERSION AA239785.1 GI:1863825

KEYWORDS
 SOURCE house mouse.
 ORGANISM Mus musculus

REFERENCE
 AUTHORS Mairra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.

TITLE The WashU-HMNI Mouse EST Project
 JOURNAL Unpublished (1996)
 CONTACT: Mairra M/Mouse EST Project
 COMMENT WashU-HMNI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810

Email: mouseest@watson.wustl.edu
 This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.lnl.gov) for further information.
 MGI:426180

Seq primer: -28ml3 rev2 ET from Amersham
 High quality sequence stop: 461.

FEATURES
 Location/Qualifiers

source

```

1..474
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:692620"
/clone_lib="Soares mouse NML"
/tissue_type="Liver"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5'; TGTATCCATCTGAGTGGAGCGGCGCGGAACTTTTCTTTTCTTTTCTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library constructed and normalized by Bento Soares and M.Fatima Bonaldo."
```

BASE COUNT 135 a 99 c 118 g 122 t
 ORIGIN

alignment_scores:
 Quality: 823.00 Length: 158
 Ratio: 5.310 Gaps: 0
 Percent Similarity: 98.101 Percent Identity: 96.203

alignment_block:
 US-09-668-482-4 x AA239785 ..

Align seg 1/1 to: AA239785 from: 1 to: 474

```

324  GlnLysValArgGluLuleuLysSerLysGlyLeuLeuCysLysSerAs 340
|||||
1  CAGAAAGTTCGAGAAAGATTAAGAGCAAGGCGCTTCTTGCAGAGCAA 50

```

```

340 nglaspasnllysleuaspmetglutlleugluinleuylsyttrileg 357
      |||||||
51 TCAAGACAAAGATTAGACATGGAACCTTTGGAACACCTTAATACACTG 100
357 lyysvaliileysgluthrleuargleuasnpropovalproglygly 373
      |||||||
101 GGTTGTCAATTAGAGAGACCCCTGGATTGAATCTCCGGTTCCAGAGGG 150
374 pheargvalalaleuylsyrphleuileuasnlytyrglinleproly 390
      |||||||
151 TTTGGGTTGCTCGAAGACCTTTGACGTGAATGATACCAAGATCCCAA 200
390 sglyttrpasnvaliletyrserilecysasprthrhisaspyalalaui 407
      |||||||
201 GGCTGGAACTGTATTTACAGTATCTGTGACACCCAGATGTGGCAGATA 250
407 lepheThrAsnlysgluThrpheasnproaspargpheserAlaProHis 423
      |||||||
251 TCTTCACTAACAGAGAGAAATTTAATCCGACCCCTTTATAGTCCCTCAT 300
424 ProgluaspalaserArgpheserPheileProphleglyglyleuar 440
      |||||||
301 CCAGAGAGTGTCTCCGGTTCAGCTTCATTCATTGGAGAGAGCCCTCG 350
440 gsercysvaliglylysgluPhealalysilleuileuylsillephethr 457
      |||||||
351 GAGCTGTGTAGCAAGAGATTGGCAAAATCTTCTTAAGATATTATACAG 400
457 algluileuAlaIArgHisCysasprThrPgluileuAsnlyProProthr 473
      |||||||
401 TGGAGCTGGTAGGCACTGTGATTGGCAGCTTCTTAATGGACTCTCTACA 450
474 MetlysrThrSerProThrValTyr 481
      |||||||
451 ATGAAGACAGAGCCCACTGTGTAC 474

seq_name: gb_est85:BF236872

seq_documentation_block:
LOCUS BF236872 545 bp mRNA EST 14-NOV-2000
DEFINITION 602027980F1 NCL_CGAP_L19 Mus musculus cDNA clone IMAGE:4163202 5',
      mRNA sequence.
ACCESSION BF236872
VERSION BF236872.1 GI:11150789
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 545)
AUTHORS NIH-MGC http://mgi.mc.ncl.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
            Email: cgaaps-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LNL at:
            http://image.lnl.gov
            Plate: LAM9446 row: p column: 19
            High quality sequence start: 2
            High quality sequence end: 545.
            Location/Qualifiers
                1..545
                /organism="Mus musculus"
                /strain="FVB/N"
                /db_xref="taxon:10090"
                /clone="IMAGE:4163202"
                /clone_lib="NCL_CGAP_L19"
FEATURES
source

```

```

/lab_host="DH10B (T1 phage-resistant)"
/notes="Organ: Liver; Vector: PCMV-SPOK6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.9 kb. Constructed by Life
Technologies. Note: this is a NCL_CGAP library."

BASE COUNT      87 a      165 c      184 g      109 t
ORIGIN

alignment_scores:
      Quality: 769.00      Length: 173
      Ratio: 4.633      Gaps: 4
Percent Similarity: 95.954      Percent Identity: 89.017

alignment_block:
US-09-668-482-4 x BF236872 ..

Align seg 1/1 to: BF236872 from: 1 to: 545

1 MetGlyLeuProAlaLeuAlaSerAlaLeuCysThrPheValLeuPr 17
|||||
35 ATGGGGCTCCCGCGCGCTGCTGGCCAGTGGCCTTGACACTTGCTGGC 84
17 OleuLeuLeuPheLeuAlaAlaileLysLeuTrpAspleuTyrCysValS 34
|||||
85 GCTGCTGCTCTTCCTGGCGCGCTCAAGCTCTGGACCTGACTGTGTA 134
34 erglyArgaspargserCysAlaLeuProleuProglyThrMetGlyS 50
|||||
135 GCAGCCCGCATGCCAGCTGGCGCTCCCTTGGCCCCCGGTACCAAGGGG 184
PheProPhe.PheGlygluThrleuGlnMetValleuGlnArgArgLysP 67
|||||
185 TTCCCATTTCTTGGGA...AACATGCAGATGTGCTTCACAGGAGAGAT 231
67 heuLeuGlnMetLysArgArgLysTyrGlyPheileTyrIlyThrHisLeu 83
|||||
232 TTTCGACAGATGAACCGCAGAAATACGCTTCATCTCAAGAGCATCTG 281
84 PheGly.ArgProThrValArgValMetGlyAlaAspAsnValArgArgI 100
|||||
282 TTGGCGCGGCC...ACGGTGGCGGTGATGGCGCGCGATTAATCTCGGCGCA 328
100 leuLeuGlyaspaspArgLeuValSerValHisTrpProAlaSerVal 116
|||||
329 TCTTGCTGGGAGAGACACCGGTTGGTGTGGGTGACACTGGCGGCTGGT 378
117 ArgThrIleleuGlyserGlyCysLeuSerAsnleuHisaspSerSerH 133
|||||
379 CGCACATCCCTGGGCGCTGCTGCCCTCCCAACCTGCACAGATTCCTGCA 428
133 slsglyArgArgLysLysValileMetArgAlaPheSerArgGluAlaLeu 150
|||||
429 CAAGCAGCGAAGAAAGATGATATGACAGCCTTCAGCGGAGGACATCC 478
150 lueyTyTyValProValileThrGluGluValGlyserSerleuGluIn 166
|||||
479 AGTGCTACGTCCCGCTGATCTGTAGGAAATGCACAGTGTCTGTGAGCAG 528
167 TrpleuSerCysGly 171
|||||
529 TGGCTAAAGCTCGGCG 543

seq_name: gb_est82:BF055367

seq_documentation_block:
LOCUS BF055367 646 bp mRNA EST 16-OCT-2000
DEFINITION 7178112.X1 Soates.NSF_F8_9W_OT_PA_P_S1 Homo sapiens cDNA clone
      IMAGE:3392591 3' similar to SW:CP26_HUMAN O43114 CTTGCHROME P450 26
      mRNA sequence.
ACCESSION BF055367
VERSION BF055367.1 GI:10809263
KEYWORDS EST.

```

SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 646)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph. D.
 Email: cgapds-remail.nih.gov
 This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.lnl.gov) for further information.
 Seq primer: -40UP from Gibco
 High quality sequence stop: 450.

FEATURES
 source
 1..646
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3392591"
 /clone.lib="Soares_NSF_F8_9W_OT_PA_P_S1"
 /lab_host="DH10B"
 /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Equal amounts of plasmid DNA from five normalized libraries were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 5 libraries. The pools consisted of the following libraries and cloneIDs: Soares NBHSP pool 1: 309384-310919, 323208-325895 Soares NB2HP pool 1: 145032-147335, 147720-148103, 148872-149255, 15002 - 150407, 151176-152327 Soares NB2HP-9W pool 1: 758280-760583, 772104-774407 Soares NBHFA pool 1: 304776-306311, 320136-322823, 326280-326653 Soares NBHOT pool 1: 723720-726407, 739080-740999 Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 202 a 129 c 120 g 194 t 1 others
 ORIGIN

alignment_scores:
 Quality: 763.00 Length: 149
 Ratio: 5.336 Gaps: 0
 Percent Similarity: 95.973 Percent Identity: 93.960

alignment_block:
 US-09-668-482-4 x BF053367/rev ..

Align seg 1/1 to reverse of: BF053367 from: 1 to: 646

```

349 11leuGlunleuLysTyrIleGlyCysValIleLysGluThrLeuAr 365
||||| ::::::::::::::::::::::::::::::::::::::::::::
646 ATTTTGGCCCACTTAATACCTTGGGTGTATTAAGAGACCCCTTNG 597
365 gleuAsnProProValProGlyGlyPheArGValAlaLeuLysThrPheG 382
||||| ::::::::::::::::::::::::::::::::::::::::::::
596 ACTGATCCCGCCAGTCCAGAGAGGTTTCGCTTCACAAACCTTTTA 547
382 IuleuAsnGlyTyrGlnIleProLysGlyTTPAsnValIleTyrSerIle 398
::: ||::::::::::::::::::::::::::::::::::::::::::
546 ATTAAATGATGATACCAAGATTCCTCAAGGCTGGAATGTATCTACAGTATC 497
399 CysAspThrHisAspValAlaGluIlePheThrAsnLysGluThrPheAs 415
||||| ::::::::::::::::::::::::::::::::::::::::::::
496 TGTGATACATCATGATGTGGCAGAGATCTTCACCAACAGAGAAATTTAA 447
415 nProAspArgPheSerAlaProHisProGluAspAlaSerArgPheSerP 432
||||| ::::::::::::::::::::::::::::::::::::::::::::
446 TCTGTACCATTCATGCTGCTCACCAGAGATGATCATCAGGTTTACGCT 397
432 heIleProheGlyGlyIleuArgSerCysValIleLysGluPheAla 448

```

```

||||| ::::::::::::::::::::::::::::::::::::::::::::
396 TCATTCATTTGGAGAGGCTTAGAGCTGTGTAGGCAAAATTTTGA 347
449 LysIleLeuLeuLysIlePheThrValGluLeuAlaArgHisCysAspTr 465
||||| ::::::::::::::::::::::::::::::::::::::::::::
346 AAAATCTTCTCAAAATATTTTACAGTGGAGCGTGGCAGCATTTGACAG 297
465 pGlnIleuAsnGlyProProThrMetLysThrSerProThrValTyrP 482
||||| ::::::::::::::::::::::::::::::::::::::::::::
296 GCAGCTCTTAATGAGACCTCTACATGAMAAACAGTCCACCGGTATC 247
482 rovalAspAsnLeuProAlaArgPheThrHisPheHisGlyGluIle 497
||||| ::::::::::::::::::::::::::::::::::::::::::::
246 CTGTGCAAAATCTCCCTCAGACATTCACCATTTCCATGGGAAATC 200

```

seq_name: gb_est43:AW174347

seq_documentation_block: 720 bp mRNA EST 16-NOV-1999
 LOCUS AW174347
 DEFINITION f142406.y1 Sugano Kawakami zebrafish DNA Danio rerio cDNA clone
 2640274 5' similar to SW:CP26_BRARE P79739 CYTOCHROME P450 26 ;,
 mRNA sequence.
 ACCESSION AW174347
 VERSION AW174347.1 GI:6440295
 KEYWORDS EST.
 SOURCE zebrafish.
 ORGANISM Danio rerio
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes; Cyprinidae; Rasbora; Danio.
 REFERENCE 1 (bases 1 to 720)
 AUTHORS Sugano S., Kawakami K., Johnson S., Li F., Marra M., Eddy S., Hillier L., Clifton S., Allen M., Gibbons M., Jost S., Kucaba T., Martin J., Pape D., Stepien M., Underwood K., Theising B., Riltter E., Bowers Y., Wylie T., Waterston R. and Wilson R.
 TITLE WashU Zebrafish EST Project 1999
 JOURNAL Unpublished (1999)
 COMMENT Other ESTs: f142406.x1
 Contact: S.L. Johnson
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 Library constructed by Dr. Sumio Sugano and Dr. Koichi Kawakami DNA Sequencing by: Washington University Genome Sequencing Center
 Seq primer: T3 ET from Amersham
 High quality sequence stop: 478.

FEATURES
 source
 1..720
 /organism="Danio rerio"
 /strain="AB"
 /db_xref="taxon:7955"
 /clone="2640274"
 /clone.lib="Sugano Kawakami zebrafish DNA"
 /sex="mixed (one male and one female, including unfertilized eggs)"
 /dev_stage="adult"
 /lab_host="DH10B (phage resistant)"
 /note="Vector: pME18S-FL3; site_1: DraIII (CACGTGNG); site_2: DraIII (CACCATGNG); 1st strand cDNA was primed with an oligo(dT) primer [ATGTGGCCCTTTTCTTTTCTTTT]; double-stranded cDNA was ligated to a DraIII adaptor [TGTGGCTTACTG], digested and cloned into distinct DraIII sites of the pME18S-FL3 vector (5' site CACTGTG, 3' site CACCATG). XhoI should be used to isolate the cDNA insert. Size selection was performed to exclude fragments <1.5kb. Library constructed by Dr. Sumio Sugano (University of Tokyo Institute of Medical Science) and kindly donated by Dr. Koichi Kawakami. Custom primers for sequencing: 5' end primer CTTCGCTTAAAGCTGCG and 3' end primer CGACCTGCAGCTCGACACA."

BASE COUNT 172 a 176 c 196 g 176 t

ORIGIN

alignment_scores: Length: 239
 Quality: 751.50
 Ratio: 3.776
 Percent Similarity: 83.264 Percent Identity: 63.180

alignment_block:
 US-09-668-482-4 x AW174347 ..

Align seg 1/1 to: AW174347 from: 1 to: 720

```

1 MetGlyLeuProAlaLeuAlaSerAlaLeuCysThrPheValLeuP 17
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
25 ATGGGGCTGTACACCTTATGTGTCACCGTGTCTCATCGTCTACC 74
17 cLeuLeuLeuPheLeuAlaAlaLeuLeuLeuTPaspLeuTyrCysValS 34
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
75 CGTTTACTCTTCTCGCGCGGTAAGTGTGGAGATGTAAATGATCC 124
34 ergLyArgAspArgSerCysAlaLeuProLeuProGlyThrMetGly 50
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
125 GACGAGTCGATCCGAACTCGAAGCTCTTACCGCGAGGTACCATGGC 174
51 PheProPhePheGlyGluThrLeuGluMetValLeuGluAlaArgGlySph 67
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
175 TTGCGGTTTCATTTGAGAAACGCTCCACGTCATCTCCAGAGAAAGATT 224
67 eLeuGluMetLysArgArgLysTyrGlyPheLeuTyrLysThrHisLeuP 84
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
225 TCTCGCATGTAAGCAAGCAATACGGGTGCATCTCAAGACGCACTCT 274
84 heGlyArgProThrValArgValMetGlyAlaAspAsnValArgArgIle 100
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
275 TCGGGAACCCACCTGTCAGGCTGATGGAGCTGATATGTAGGACAGATT 324
101 LeuLeuGlyAspAspArgLeuValSerValHisTrrProAlaSerValAr 117
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
325 CTGCTGGCGGAACACAAAGCTGTCTCTTCATGCGCCAGCATAGTGGAG 374
117 gThrIleLeuGlySerGlyCysLeuSerAsnLeuHisAspSerSerHisL 134
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
375 AACCATCTCGGCTCTGACACCTCTCCATGTCATGAGAGTTCAACACA 424
134 ySGlnArgLysLysValIleMetArgAlaPheSerArgGluAlaLeuGlu 150
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
425 AAAACAAGAAAAAGCCATTATGAGGGCTCTCTCGAGATGCTCTGAG 474
151 CysTyrValProValIleThrGluGluValGlySerSerLeuGluGlnTr 167
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
475 CACTCATCTCCCGTGCATCCATGAGAGTGAAGACCGCCATACAGAGATTG 524
167 Pleu.....SerCysGlyGluArgGlyLeuLeuValTyrProGluV 181
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
525 GCTGCAAAAAGACTCTCTGC.....CTGCTGTTTATTCAGATTA 562
181 alLyArgLeuMetPheArgIleAlaMetArgIleLeuLeuGlyCysGlu 197
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
563 TGAAGATACTCATGTTGCGGATAGCTATGAGAAATCCTGCTGGTGA 612
198 Pro...GlnLeuAlaGlyAspGlyAspSerGluGlnGlnLeuValGluAl 213
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
613 CCACAGCAAAATATAGACTGAC.....GAGCATGAGAGCTGTGGAAGC 653
213 apheGluGluMetThrArgAsnLeuPheSerLeuProIleAspValProp 230
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
654 TTGGAGCCAAATGATCAGGAGCTGTCTGCTGCCAATCCACGTTCTTT 703
230 heSerGlyLeuTyrArg 235
||| |||||||
704 TCATGTGCTGTACAGG 720

```

seq_name: gb_est51:AW765767

seq_documentation_block:

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

seq_name: gb_est51:AW765767

seq_documentation_block:

LOCUS 669 bp mRNA EST 16-FEB-2001

DEFINITION da77a02.y1 Harland stage 19-23 *Xenopus laevis* cDNA clone

IMAGE:3200906 5' similar to gb:gb1AF057566.11AF057566 *Xenopus laevis* retinoic acid converting enzyme (XENOPUS);, mRNA sequence.

ACCESSION AW765767

VERSION AW765767.1 GI:7697743

KEYWORDS EST.

SOURCE African clawed frog.

ORGANISM *Xenopus laevis*

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae; Xenopodinae; *Xenopus*.

AUTHORS 1 (bases 1 to 669)

 Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L., Pape,D., Martin,J., Wylie,T., Underwood,K., Theising,B., Bowers,Y., Person,B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R., Waterston,R. and Wilson,R.

TITLE Washu *Xenopus* EST project, 1999

JOURNAL Unpublished (1999)

COMMENT Other_ESTs: da77a02.x1

 Contact: Sandy Clifton, Ph.D.

 Washu *Xenopus* EST project, 1999

 Washington University School of Medicine

 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

 Tel: 314 286 1800

 Fax: 314 286 1810

 Email: est@watson.wustl.edu

 Library constructed by R. Harland, PhD, (University of California, Berkeley)

 DNA Sequencing by: Washington University Genome Sequencing Center

 Clone distribution: *Xenopus* clones from this library are available through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov/image/html/tresources.shtml

 Seq primer: -40RP from Gibco

 High quality sequence stop: 495.

FEATURES

 source

 location/Qualifiers

 1..669

 /organism="Xenopus laevis"

 /db_xref="taxon:8355"

 /clone IMAGE:3200906"

 /clone_lib="Harland stage 19-23"

 /tissue_type="neura"

 /dev_stage="stage 19-23"

 /lab_host="DH10B (phage-resistant)"

 /note="vector: PCS107 (custom); Site_1: NotI; Site_2: SalI; cDNA made by oligo-dT priming. Library constructed by Dr. Francesca Mariani in the laboratory of R. Harland, Ph.D. (University of California, Berkeley). References: XBF-2 is a transcriptional repressor that converts ectoderm into neural tissue. Mariani, FV, Harland, RM, Development. 1998 Dec;125(24):5019-31. PMID: 9811586; UI: 99030283; Use of large-scale expression cloning screens in the *Xenopus laevis* tadpole to identify gene function. Grammer TC, Liu KJ, Mariani FV, Harland RM, Dev Biol. 2000 Dec 15;228(2):197-210. PMID: 11112324; UI: 20564075; Note: This is a *Xenopus* Gene Collection (XGC) library."

BASE COUNT

137 a 206 c 187 g 139 t

ORIGIN

alignment_scores: Length: 199
 Quality: 731.50
 Ratio: 4.228
 Percent Similarity: 86.935 Percent Identity: 71.357

alignment_block:
 US-09-668-482-4 x AW765767 ..

Align seg 1/1 to: AW765767 from: 1 to: 669

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2001, 13:27:05 ; Search time 88.14 Seconds
(without alignments)
126.891 Million cell updates/sec

Title: US-09-668-482-32

Perfect score: 2612
Sequence: 1 MGIPALLASALCTFVLPILL.....PTVYPVDNLPARFTYFGDI 497

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA: *
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep.*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep.*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep.*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep.*
5: /cgn2_6/ptodata/2/1aa/PTUS.COMB.pep.*
6: /cgn2_6/ptodata/2/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2612	100.0	497	4	US-08-882-164D-32 Sequence 32, Appl
2	2437	93.3	497	3	US-08-724-466B-4 Sequence 4, Appl
3	2437	93.3	497	4	US-08-882-164D-4 Sequence 4, Appl
4	1734	66.4	492	3	US-08-724-466B-2 Sequence 2, Appl
5	1734	66.4	492	4	US-08-882-164D-2 Sequence 2, Appl
6	450	17.2	472	2	US-08-622-166A-2 Sequence 2, Appl
7	450	17.2	472	2	US-08-622-166A-4 Sequence 4, Appl
8	358.5	13.7	511	4	US-08-991-677-4 Sequence 9, Appl
9	337	12.9	520	2	US-09-091-432-2 Sequence 2, Appl
10	300	11.5	576	3	US-08-948-564-16 Sequence 16, Appl
11	284.5	10.9	510	3	US-08-948-564-4 Sequence 4, Appl
12	270.5	10.4	496	1	US-08-313-075A-50 Sequence 50, Appl
13	262.5	10.0	500	4	US-09-292-768-68 Sequence 68, Appl
14	261.5	10.0	500	4	US-09-292-768-4 Sequence 4, Appl
15	261.5	10.0	500	4	US-09-292-768-70 Sequence 70, Appl
16	258.5	9.9	476	1	US-08-313-075A-10 Sequence 30, Appl
17	256.5	9.8	500	3	US-08-881-784-9 Sequence 9, Appl
18	255.5	9.8	513	3	US-08-948-564-2 Sequence 2, Appl
19	255.5	9.8	513	3	US-08-948-564-6 Sequence 6, Appl
20	249	9.5	510	3	US-08-606-505B-66 Sequence 66, Appl
21	249	9.5	510	4	US-09-616-990-66 Sequence 66, Appl
22	242	9.3	500	2	US-08-314-601-2 Sequence 2, Appl
23	242	9.3	500	5	PCT-US95-13051-2 Sequence 2, Appl
24	240	9.2	504	1	US-08-457-274A-25 Sequence 25, Appl
25	240	9.2	504	1	PCT-US95-05758-25 Sequence 25, Appl
26	239	9.2	496	4	US-09-292-768-64 Sequence 64, Appl
27	238.5	9.1	509	3	US-08-948-564-18 Sequence 18, Appl

28	237.5	9.1	490	1	US-08-201-118-3 Sequence 3, Appl
29	237.5	9.1	490	1	US-08-201-118-9 Sequence 9, Appl
30	237.5	9.1	490	2	US-08-238-821B-3 Sequence 3, Appl
31	237.5	9.1	490	2	US-08-238-821B-9 Sequence 9, Appl
32	237.5	9.1	490	5	PCT-US95-05744-3 Sequence 3, Appl
33	237.5	9.1	490	5	PCT-US95-05744-9 Sequence 9, Appl
34	237.5	9.1	532	3	US-08-948-564-10 Sequence 10, Appl
35	235	9.0	496	3	US-08-881-784-1 Sequence 1, Appl
36	235	9.0	496	4	US-09-292-768-2 Sequence 2, Appl
37	235	9.0	496	4	US-09-292-768-66 Sequence 66, Appl
38	234.5	9.0	496	4	US-09-172-339-6 Sequence 6, Appl
39	231.5	8.9	490	1	US-08-201-118-1 Sequence 1, Appl
40	231.5	8.9	490	2	US-08-238-821B-1 Sequence 1, Appl
41	231.5	8.9	490	5	PCT-US95-05744-1 Sequence 1, Appl
42	231	8.8	490	1	US-08-201-118-13 Sequence 13, Appl
43	231	8.8	490	2	US-08-238-821B-13 Sequence 13, Appl
44	231	8.8	490	5	PCT-US95-05744-13 Sequence 13, Appl
45	228.5	8.7	517	1	US-08-457-274A-2 Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-08-882-164D-32
Sequence 32, Application US/08882164D
Patent No. 6306624
GENERAL INFORMATION:
APPLICANT: Petkovich, P. Martin, White, Jay A.,
TITLE OF INVENTION: Retinoid Metabolizing Protein
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: Blake, Cassels & Graydon
STREET: Box 25, Commerce Court West
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5L 1A9
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 Mb storage
COMPUTER: COMPAG, IBM PC compatible
OPERATING SYSTEM: MS-DOS 5.1
SOFTWARE: WORD PERFECT
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/882,164D
FILING DATE: June 25, 1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/667,546
FILING DATE: June 21, 1996
APPLICATION NUMBER: 08/724,466
FILING DATE: October 1, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Hunt, John C.
REGISTRATION NUMBER: 36,424
REFERENCE/DOCKET NUMBER: 50767/00010
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 863-4344
TELEFAX: (416) 863-2653
INFORMATION FOR SEQ ID NO: 32:
SEQUENCE CHARACTERISTICS:
LENGTH: 497 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-882-164D-32

Query Match 100.0%; Score 2612; DB 4; Length 497;
Best Local Similarity 100.0%; Pred. No. 5.5e-276;
Matches 497; Conservative 0; Mismatches 0; Gaps 0;
QY 1 MGIPALLASALCTFVLPILLFLAALKMDICVSSRDSKALPLPPGTWGPFFGTGLWM 60

```

Db 1 MGLPALLASALCTEVLPLLLFLAALKIMDLVCSRRDSCALPLPPTMGPFPEFTIOM 60
QY 61 VLORRRFLQMRKRYGFIYKTHLGRPTVRMGADNVRILLGHRILVSHMPASVFTIL 120
Db 61 VLORRRFLQMRKRYGFIYKTHLGRPTVRMGADNVRILLGHRILVSHMPASVFTIL 120
QY 121 GAGCLSNLHDSHQRKRVIMQAFSREALQCYVIVIAEVSCLQOMLSCGERGLVYPE 180
Db 121 GAGCLSNLHDSHQRKRVIMQAFSREALQCYVIVIAEVSCLQOMLSCGERGLVYPE 180
QY 181 VKRLMFRIAMRILLGCEGPGAGGEDQOLVFAFEEMTRNLFSLPIDVPFSGLYRGVKAR 240
Db 181 VKRLMFRIAMRILLGCEGPGAGGEDQOLVFAFEEMTRNLFSLPIDVPFSGLYRGVKAR 240
QY 241 NLHARIEENIRAKIRRLQATEPDGCKDALQILLIHSWEMGERLDMQALQKOSTELLFG 300
Db 241 NLHARIEENIRAKIRRLQATEPDGCKDALQILLIHSWEMGERLDMQALQKOSTELLFG 300
QY 301 GHETTASAASTSLTYLGLYPHVLOKVREELKSKGLCKSNODNKLDMETLEQLKYICVI 360
Db 301 GHETTASAASTSLTYLGLYPHVLOKVREELKSKGLCKSNODNKLDMETLEQLKYICVI 360
QY 361 KETRLRMPVPVGGFRVALKTFEELNGYQIPKGMNYYISICDTHVADIFTNKEEFPDREI 420
Db 361 KETRLRMPVPVGGFRVALKTFEELNGYQIPKGMNYYISICDTHVADIFTNKEEFPDREI 420
QY 421 VPHEDASRSFIFPGGGLSCVCKEFAKILKIFTVELARHCDMOLLNPGPTMKTSPV 480
Db 421 VPHEDASRSFIFPGGGLSCVCKEFAKILKIFTVELARHCDMOLLNPGPTMKTSPV 480
QY 481 YPVDNLPAFTFYOGDI 497
Db 481 YPVDNLPAFTFYOGDI 497

```

RESULT 2

US-08-724-466B-4
Sequence 4, Application US/08724466B
Patent No. 6063606

GENERAL INFORMATION:

APPLICANT: Petkovich, P. Martin, White, Jay A.,
APPLICANT: Beckett, Barbara R., Jones, Glenville
TITLE OF INVENTION: Retinoid Metabolizing Protein
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:

ADDRESSEE: Blake, Cassels & Graydon
STREET: Box 25, Commerce Court West
CITY: Toronto
ZIP: M5L 1A9

COUNTRY: Canada

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 MB storage
COMPUTER: COMPAQ, IBM PC compatible
OPERATING SYSTEM: MS-DOS 5.1
SOFTWARE: WORD PERFECT

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/724,466B
FILING DATE: October 1, 1996

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/667,546
FILING DATE: June 21, 1996

ATTORNEY/AGENT INFORMATION:

NAME: Hunt, John C.
REGISTRATION NUMBER: 36,424
REFERENCE/DOCKET NUMBER: 50767/00004
TELECOMMUNICATION INFORMATION:

TELEPHONE: (416) 863-4344
TELEFAX: (416) 863-2653

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:
LENGTH: 497 amino acids
TYPE: amino acid

STRANDEDNESS: single
TOPOLOGY: linear
US-08-724-466B-4

Query Match 93.3%; Score 2437; DB 3; Length 497;
Best Local Similarity 93.4%; Pred. No. 6.7e-257;
Matches 464; Conservative 14; Mismatches 19; Indels 0; Gaps 0;

```

QY 1 MGLPALLASALCTEVLPLLLFLAALKIMDLVCSRRDSCALPLPPTMGPFPEFTIOM 60
Db 1 MGLPALLASALCTEVLPLLLFLAALKIMDLVCSRRDSCALPLPPTMGPFPEFTIOM 60
QY 61 VLORRRFLQMRKRYGFIYKTHLGRPTVRMGADNVRILLGHRILVSHMPASVFTIL 120
Db 61 VLORRRFLQMRKRYGFIYKTHLGRPTVRMGADNVRILLGHRILVSHMPASVFTIL 120
QY 121 GAGCLSNLHDSHQRKRVIMQAFSREALQCYVIVIAEVSCLQOMLSCGERGLVYPE 180
Db 121 GAGCLSNLHDSHQRKRVIMQAFSREALQCYVIVIAEVSCLQOMLSCGERGLVYPE 180
QY 181 VKRLMFRIAMRILLGCEGPGAGGEDQOLVFAFEEMTRNLFSLPIDVPFSGLYRGVKAR 240
Db 181 VKRLMFRIAMRILLGCEGPGAGGEDQOLVFAFEEMTRNLFSLPIDVPFSGLYRGVKAR 240
QY 241 NLHARIEENIRAKIRRLQATEPDGCKDALQILLIHSWEMGERLDMQALQKOSTELLFG 300
Db 241 NLHARIEENIRAKIRRLQATEPDGCKDALQILLIHSWEMGERLDMQALQKOSTELLFG 300
QY 301 GHETTASAASTSLTYLGLYPHVLOKVREELKSKGLCKSNODNKLDMETLEQLKYICVI 360
Db 301 GHETTASAASTSLTYLGLYPHVLOKVREELKSKGLCKSNODNKLDMETLEQLKYICVI 360
QY 361 KETRLRMPVPVGGFRVALKTFEELNGYQIPKGMNYYISICDTHVADIFTNKEEFPDREI 420
Db 361 KETRLRMPVPVGGFRVALKTFEELNGYQIPKGMNYYISICDTHVADIFTNKEEFPDREI 420
QY 421 VPHEDASRSFIFPGGGLSCVCKEFAKILKIFTVELARHCDMOLLNPGPTMKTSPV 480
Db 421 VPHEDASRSFIFPGGGLSCVCKEFAKILKIFTVELARHCDMOLLNPGPTMKTSPV 480
QY 481 YPVDNLPAFTFYOGDI 497
Db 481 YPVDNLPAFTFYOGDI 497

```

RESULT 3

US-08-882-164D-4
Sequence 4, Application US/08882164D
Patent No. 6306624

GENERAL INFORMATION:

APPLICANT: Petkovich, P. Martin, White, Jay A.,
APPLICANT: Beckett, Barbara R., Jones, Glenville
TITLE OF INVENTION: Retinoid Metabolizing Protein
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:

ADDRESSEE: Blake, Cassels & Graydon
STREET: Box 25, Commerce Court West
CITY: Toronto
STATE: Ontario

COUNTRY: Canada

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 MB storage
COMPUTER: COMPAQ, IBM PC compatible
OPERATING SYSTEM: MS-DOS 5.1
SOFTWARE: WORD PERFECT

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/882,164D

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/667,546
FILING DATE: June 21, 1996

```

APPLICATION NUMBER: 08/724,466
FILING DATE: October 1, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Hunt, John C.
REGISTRATION NUMBER: 36,424
REFERENCE/DOCKET NUMBER: 50767/00010
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 863-4344
TELEFAX: (416) 863-2653
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 497 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-882-16AD-4

Query Match          93.3%; Score 2437; DB 4; Length 497;
Best Local Similarity 93.4%; Pred. No. 6.7e-257;
Matches 464; Conservative 14; Mismatches 19; Indels 0; Gaps 0;

QY 1 MGIPALLASALCTFVPLPILFLAALKIMDLVCVSSDRSCALPLPGTMGFPEFTIOM 60
DB 1 MGIPALLASALCTFVPLPILFLAALKIMDLVCVSGDRSCALPLPGTMGFPEFTIOM 60
QY 61 VLORRKFLOMKRRKRYGIYKTHLFGRTVYVGMADNVRRILGHRILVSVHPASVRTIL 120
DB 61 VLORRKFLOMKRRKRYGIYKTHLFGRTVYVGMADNVRRILGHRILVSVHPASVRTIL 120
QY 121 GAGCLSNLHDSHKKRKRYIMQAFSREALQCYLVYAEVSSCLEBWLSCGEGGLLYPE 180
DB 121 GAGCLSNLHDSHKKRKRYIMQAFSREALQCYLVYAEVSSCLEBWLSCGEGGLLYPE 180
QY 181 VKRIMFRIAMRILGCEPPAGGEGDEQOLVEAFEEFTRILFSLPIDVPSGLYRGVKAR 240
DB 181 VKRIMFRIAMRILGCEPPAGGEGDEQOLVEAFEEFTRILFSLPIDVPSGLYRGVKAR 240
QY 241 NLIHARIEENIRAKIRRLQATEPDGCKDALQLLIEHSWGERLDMQALKOSTELLFG 300
DB 241 NLIHARIEENIRAKIRRLQATEPDGCKDALQLLIEHSWGERLDMQALKOSTELLFG 300
QY 301 GHETTASATSLTYLGLYHYVLOKVEEIKSKGLCKSNODNKLDMETLEOLKTYGCVI 360
DB 301 GHETTASATSLTYLGLYHYVLOKVEEIKSKGLCKSNODNKLDMETLEOLKTYGCVI 360
QY 361 KETRLNRPVYVGGFRVALKTFELNGYQIPKGMNVIYSICDTHVADVFITNKEEFPDRFI 420
DB 361 KETRLNRPVYVGGFRVALKTFELNGYQIPKGMNVIYSICDTHVADVFITNKEEFPDRFI 420
QY 421 VPHPEDASRFSFIPIFGGGLRSCVGEKPAKILKIFTYELARHCDMOLLNPPYTKTSPTV 480
DB 421 VPHPEDASRFSFIPIFGGGLRSCVGEKPAKILKIFTYELARHCDMOLLNPPYTKTSPTV 480
QY 481 YPVNDLPARFTYFOGDI 497
DB 481 YPVNDLPARFTYFOGDI 497

RESULT 4
US-08-724-466B-2
Sequence 2, Application US/08724466B
GENERAL INFORMATION:
APPLICANT: Petkovich, P. Martin, White, Jay A.,
APPLICANT: Beckett, Barbara R., Jones, Glenville
TITLE OF INVENTION: Retinoid Metabolizing Protein
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESSEE: Blake, Cassels & Graydon
STREET: Box 25, Commerce Court West
CITY: Toronto
ZIP: M5L 1A9

```

```

COUNTRY: Canada
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 Mb storage
COMPUTER: COMPAQ, IBM PC compatible
OPERATING SYSTEM: MS-DOS 5.1
SOFTWARE: WORD PERFECT
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/724,466B
FILING DATE: October 1, 1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/667,546
FILING DATE: June 21, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Hunt, John C.
REGISTRATION NUMBER: 36,424
REFERENCE/DOCKET NUMBER: 50767/00004
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 863-4344
TELEFAX: (416) 863-2653
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 492 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-724-466B-2

```

```

Query Match          66.4%; Score 1734; DB 3; Length 492;
Best Local Similarity 67.7%; Pred. No. 2.9e-180;
Matches 336; Conservative 61; Mismatches 89; Indels 10; Gaps 4;

```

```

QY 1 MGIPALLASALCTFVPLPILFLAALKIMDLVCVSSDRSCALPLPGTMGFPEFTIOM 60
DB 1 MGIPALLASALCTFVPLPILFLAALKIMDLVCVSSDRSCALPLPGTMGFPEFTIOM 60
QY 61 VLORRKFLOMKRRKRYGIYKTHLFGRTVYVGMADNVRRILGHRILVSVHPASVRTIL 120
DB 61 VLORRKFLOMKRRKRYGIYKTHLFGRTVYVGMADNVRRILGHRILVSVHPASVRTIL 120
QY 121 GAGCLSNLHDSHKKRKRYIMQAFSREALQCYLVYAEVSSCLEBWLSCGEGGLLYPE 177
DB 121 GAGCLSNLHDSHKKRKRYIMQAFSREALQCYLVYAEVSSCLEBWLSCGEGGLLYPE 177
QY 178 YPEVKLMPRIAMRILGCEPPAGGEGDEQOLVEAFEEFTRILFSLPIDVPSGLYRGV 237
DB 178 YPEVKLMPRIAMRILGCEPPAGGEGDEQOLVEAFEEFTRILFSLPIDVPSGLYRGV 237
QY 237 KANLILHARIEENIRAKIRRLQATEPDGCKDALQLLIEHSWGERLDMQALKOSTELL 297
DB 237 KANLILHARIEENIRAKIRRLQATEPDGCKDALQLLIEHSWGERLDMQALKOSTELL 297
QY 298 LFGHETTASATSLTYLGLYHYVLOKVEEIKSKGLCKSNODNKLDMETLEOLKTYG 357
DB 298 LFGHETTASATSLTYLGLYHYVLOKVEEIKSKGLCKSNODNKLDMETLEOLKTYG 357
QY 358 CVIKETRLNRPVYVGGFRVALKTFELNGYQIPKGMNVIYSICDTHVADVFITNKEEFPDR 417
DB 358 CVIKETRLNRPVYVGGFRVALKTFELNGYQIPKGMNVIYSICDTHVADVFITNKEEFPDR 417
QY 418 RFIYPHEDASRFSFIPIFGGGLRSCVGEKPAKILKIFTYELARHCDMOLLNPPYTKTS 477
DB 418 RFIYPHEDASRFSFIPIFGGGLRSCVGEKPAKILKIFTYELARHCDMOLLNPPYTKTS 477
QY 478 PTYVPNDLPARFTYF 493
DB 478 PTYVPNDLPARFTYF 493

```

```

RESULT 5
US-08-882-16AD-2
Sequence 2, Application US/0888216AD
Patent No. 6306624

```

```

GENERAL INFORMATION:
APPLICANT: Pelkovich, P. Martin, White, Jay A.,
APPLICANT: Beckett, Barbara R., Jones, Glenville
TITLE OF INVENTION: Retinoid Metabolizing Protein
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESS: Blake, Cassels & Graydon
STREET: Box 25, Commerce Court West
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5L 1A9

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 Mb storage
COMPUTER: COMPAQ, IBM PC compatible
OPERATING SYSTEM: MS-DOS 5.1
SOFTWARE: WORD PERFECT
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/882,164D
FILING DATE: June 25, 1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/667,546
FILING DATE: June 21, 1996
APPLICATION NUMBER: 08/724,466
FILING DATE: October 1, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Hunt, John C.
REGISTRATION NUMBER: 36,424
REFERENCE/DOCKET NUMBER: 50767/00010
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 863-4344
TELEFAX: (416) 863-2653
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 492 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-882-164D-2

```

```

Query Match          66.4%; Score 1734; DB 4; Length 492;
Best Local Similarity 67.7%; Pred. No. 2.9e-180;
Matches 336; Conservative 61; Mismatches 89; Indels 10; Gaps 4;

QY 1 MGLPLLASALCTEVLPLILFLAALKLMDLYCVSSNRSCALPLPGTMGPFPGFTLQM 60
   ||| | : : ||| ||| ||| ||| : : : | : ||| ||| ||| |||
Db 1 MGLTYIMVTFCTIVLPVLLFAVAKWEMLMIRVDPCRSPLPGTMGLPFIETLQD 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 61 VLQRRKFLQMKRRKRGFTYKTHLFGRPYRVAGADNVRRIILGHRLYSVHMPASVRTL 120
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61 ILQRRKFLQMKRQKIGCYKTHLFGNPYRVAGADNVRRIILGHRLYSVQMPASVRTL 120
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 121 GAGCISLNHDSHROKRVIMQAFREALQCYLVLYAEVSSCLEQWL---SCGERGLY 177
   ||| | : : ||| | : : ||| ||| ||| : : ||| ||| ||| |||
Db 121 GSDTISNVHGVQHKKKRAIMKRAFSRDALHETIPYIQGEVKAIDEMLOKDC----VLV 176
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 178 YPEVKRIEMFRIAMRIILGCEPGRAGGEDQOLVAFEBMTRNLSPLIDVPFSGLYRGV 237
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 177 YPEMKKLMFRIAMRIILGFEPEQI--KTDEQLVFAFEMIKNLSPLIDVPFSGLYRGL 234
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 238 KARNLIHARIENTIRAKRRIRLOATEPDGCKDALQTLTSHSWMERGERDMDALKSSREL 297
   ||| ||| ||| ||| ||| | : : ||| ||| ||| | : : ||| ||| |||
Db 235 RARNPIHSHKIENTIRKKIQD--DDNENBQKYDADQLLHNSRRSDPEPSLOMKKEATL 293
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 298 LFGGHETTASATSLITVGLYPHVLOKVRREIRKSKGLCKSNODNKLDMETLEQLYIG 357
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 294 LFGGHETTASATSLITVGLYPHVLOKVRREVEQKVMGMVYTPKGLSMELDLQKTYG 353
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 358 CVIETETRLNPPVGGFVAATKTELNQYQIPKGMNVIYSICDHDVADITTNKEEFPD 417
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 354 CVIETETRLNPPVGGFVAATKTELNQYQIPKGMNVIYSICDHDVADVPNKEEFOPE 413

```

```

QY 418 RFIYHPEDASRFSPFGGRLRSCVGRKPAKILKTFVFLARHCMDLNGPPYMKTS 477
   ||| | : : ||| | : : ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 414 RFMSKGLSDGRFENYIFPGGSRMCVGRKPAKILKTFVFLARHCMDLNGPPYMKTG 473
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 478 PTVPVDNLPAFETVF 493
   ||| ||| ||| |||
Db 474 PTVPVDNLPTKFTSY 489

```

```

RESULT 6
US-08-622-166A-2
Sequence 2, Application US/08622166A
Patent No. 5952545
GENERAL INFORMATION:
APPLICANT: KONCZ, CSABA
APPLICANT: MATHUR, JAIDEEP
APPLICANT: SZEKERES, MIKLOS
APPLICANT: ALTMANN, THOMAS
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING CYTOCHROME
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESS: Birch, Stewart, Kolasch & Birch, LLP
STREET: P.O. Box 747
CITY: Falls Church
STATE: Virginia
COUNTRY: USA
ZIP: 22040-0747
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/622,166A
FILING DATE: 27-Mar-1996
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: SVENSSON, LEONARD R.
REGISTRATION NUMBER: 30,330
REFERENCE/DOCKET NUMBER: 0147-0153P
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 205-8000
TELEFAX: (703) 205-8050
TELEX: 248345
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 472 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-622-166A-2

```

```

Query Match          17.2%; Score 450; DB 2; Length 472;
Best Local Similarity 26.6%; Pred. No. 2.8e-40;
Matches 134; Conservative 90; Mismatches 217; Indels 62; Gaps 15;

QY 6 ILASALCTFVPLILFLAALKLMDLYCVSSNRSCALPLPGTMGPFPGFTLQAV---- 61
   ||| | : : ||| | : : ||| ||| ||| ||| ||| ||| ||| |||
Db 9 ILSTIAAGFLILL-----RRTYRRMKGLPESLPLPLGTEFFOLIGAVK 52
   ||| | : : ||| | : : ||| ||| ||| ||| ||| ||| ||| |||
QY 62 -LQRRKFLQMKRRKRGFTYKTHLFGRPYRVAGADNVRRIILGHRLYSVHMPASVRTL 120
   ||| | : : ||| | : : ||| ||| ||| ||| ||| ||| ||| |||
Db 53 TENPEPFI DERVARYAGVFMTHLFGEPITFSADPTNRFVLONECKLRECSYPAICML 112
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 121 GAGCISLNHDSHROKRVIMQAFREALQCYLVLYAEVSSCLEQWLSCGERGLYRV 179
   ||| | : : ||| | : : ||| ||| ||| ||| ||| ||| ||| |||
Db 113 GHSLSLKKSGSLHKKRMSHLSLTFANSSIIKDHMLMDIDRLVRFNIDSWSS---RVLLME 168
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 180 EVKRIEMFRIAMRIILGCEPGRAGGEDQOLVAFEBMTRNLSPLIDVPFSGLYRGV 238
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 169 EAKITTFELTYKQLMSFPD-----GEWSESLRKEYLIVYBGFSLDPL-PSYYTKRAIQ 222

```


QY 172 ERLGLVYEVKRLMRLAMRILGCEPGAGGEDEQLVEAFEMTR--NLFSLPIDVP 229
 Db 176 E---LVRLTRKNTTYRAFGTI-----SHEDQDFVAILDEFSOLFSAFNADP 223
 QY 230 F-SGLYGV-----KARNLHARIEENTIRAKIR--LOATEPDGCKDALOLLIEHSWR 281
 Db 224 WLKVVPOGINVRLNKKARGALDGFIDKIIDHIOGSKNSEVDTDMVDLLAFGEAKV 283
 QY 282 GERIDMO-ALK-----QSSTELLFGGHETTASATSLITYLGLYPHYLQVREIKSG 334
 Db 284 SESDDLNSTIKLTDONIKALMDVFGTEYASAIEMAMTELKSPEDLKRVQOE----- 338
 QY 335 LCKSNODNKLDMETLEBOLKXIGCVIKETRLNPNVPGGFVALKTEFLNGYOIPKGMV 394
 Db 339 LAVVYGLDRVREKDEKLTLYLCKVLEKVRRLHPPIPLLHETADAEVSGYIIPAKSRV 398
 QY 395 IYSCIDHVDADITNKEENPDREIVPHPED--ASRPSIFPGGGLRSCVKEFAKILL 452
 Db 399 MINACAIGRDKNSWADPTFRPSRFLKDGVPDFKGNFEEIFPGSGRRSCPGMGLIYAL 458
 QY 453 KIFVELARHC-DMQLNGPPTMKT 477
 Db 459 ET-TVAHLHCFWEL---PDGMKPS 480

RESULT 9

US-09-091-432-2
 Sequence 2, Application US/09091432
 Patent No. 5981837

GENERAL INFORMATION:

APPLICANT: Chapple, Clint
 TITLE OF INVENTION: A Method For Regulation Of Plant Lignin Composition
 FILE REFERENCE: 7024-325
 CURRENT APPLICATION NUMBER: US/09/091,432
 EARLIER FILING DATE: 1998-06-18
 EARLIER APPLICATION NUMBER: PCT/US96/20094
 EARLIER FILING DATE: 1996-12-19
 EARLIER APPLICATION NUMBER: US 60/009,119
 EARLIER FILING DATE: 1995-12-22
 EARLIER APPLICATION NUMBER: US 60/013,388
 EARLIER FILING DATE: 1996-03-14
 NUMBER OF SEQ ID NOS: 3
 SOFTWARE: Microsoft Word 2.0C
 SEQ ID NO 2
 LENGTH: 520
 TYPE: PRT
 ORGANISM: Artificial
 FEATURE:
 OTHER INFORMATION: Sequence is deduced from DNA sequence of SEQ ID NO:1
 Patent No. 5981837
 US-09-091-432-2

Query Match 12.9%; Score 337; DB 2; Length 520;
 Best Local Similarity 25.8%; Pred. No. 6,8e-28;
 Matches 133; Conservative 90; Mismatches 194; Indels 98; Gaps 23;
 QY 15 VLPLLLALAKLMDLYCYSSRDSCALPLPPTMGFFPFGETLQWYLQRRKFLQMKRRK 74
 Db 22 VVSLFLFIS-----FTRRRRP---PYPPRPGMPTIIGNMLMDQLTHRLANLAKK 70
 QY 75 YGFIYKTHL-FGRPIYRVGADNVRRIILGEHRLVSVMHMDASVRTLLGACLSNL-HDSS 132
 Db 71 YGGL--CHLRMGFLHMYAVSSPEVARQVLOVDVSFNSRPATL-----AISLYLYDRA 121
 QY 133 -----HKOKKVIQMOAFSREALQCYLV-----IAEVSQCLQWMTSCGERGL 175
 Db 122 DNAPAHYGFPMQMRKVYKVFSSRAESWASVRDEVDMKMSVSCNKGKPIVNGEQ-- 179
 QY 176 LVYPEVKRLMFLRIAMRILGCEPGAGGEDEQLVEAFEMTR--NLFSLPIDVPFSGL 233
 Db 180 -IFALTRNTTYAARG--SACEKG-----QDEFIRIQLQFSKLFCAINVADFTYFPGW 229

QY 224 Y-----RGVKARNLHARIEENT--RAKTRLOATEPDG-----CKDALOLLIEHS 278
 Db 230 IDPOGINRKYARKNDLDFIDDIIDEHMKRKENONAVIDDDVDVDTMDVDDLAFYSEEA 289
 QY 279 WERGRLDMQ-----ALKOSSTELLFGGHETTASATSLITYLGLYPHYLQVREI 330
 Db 290 KLVSETADLQNSIKLTRNINIKAIIMDVFGTEYASAIEMALIELLSPEDLKRVOEL 349
 QY 331 -KSKGLCKSNODNKLDMETLEBOLKXIGCVIKETRLNPNVPGGFVALKTEFLNGYOIP 389
 Db 350 AEVYGL-----DRVEESDEIKLTYLCKLCTLRMPHPPIPLLHETADETSIDGFIP 403
 QY 390 KGMVYIYSCIDHVDADITNKEENPDREIVPHPED--ASRPSIFPGGGLRSCVKEF 447
 Db 404 KSKRMIAEAFIAGDPDPTFRPSRFLKDGVPDFKGNFEEIFPGSGRRSCPGMO- 462
 QY 448 AKILKIFVELA---RHC-DMQLNGPPTMKT 477
 Db 463 ---LGLYALDLAVAHILHCFWEL---PDGMKPS 490

RESULT 10

US-08-948-564-16
 Sequence 16, Application US/08948564
 Patent No. 6121512

GENERAL INFORMATION:

APPLICANT: Siminszky, Balazs
 APPLICANT: Dewey, Ralph E.
 TITLE OF INVENTION: No. 6121512el Cytochrome P-450 Constructs and
 TITLE OF INVENTION: Methods of Producing Herbicide-Resistant Transgenic Plants
 NUMBER OF SEQUENCES: 23
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Virginia C. Bennett
 STREET: PO Box 37428
 CITY: Raleigh
 STATE: No. 6121512ch Carolina
 COUNTRY: USA
 ZIP: 27627
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/948,564
 FILING DATE:
 CLASSIFICATION: 800
 ATTORNEY/AGENT INFORMATION:
 NAME: Bennett, Virginia C.
 REGISTRATION NUMBER: 37,092
 REFERENCE/DOCKET NUMBER: 5051-409
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 919-854-1400
 TELEFAX: 919-854-1401
 INFORMATION FOR SEQ ID NO: 16:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 576 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-948-564-16

Query Match 11.5%; Score 300; DB 3; Length 576;
 Best Local Similarity 23.7%; Pred. No. 8.7e-24;
 Matches 128; Conservative 86; Mismatches 187; Indels 138; Gaps 21;
 QY 41 ALPLPPT-----MGPFEGETLQWYLQRRKFLQMKRRKYGFIYKTHLFGSPYRVGADN 96
 Db 78 SMPIARGAVSGLGRPLFTSLIDWLE-----HGAIVYKL-AFGKAFVAVSDPI 125


```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: AU PCT/AU93/00127
; FILING DATE: 25-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Digiglio, Frank S.
; REGISTRATION NUMBER: 31,346
; REFERENCE/DOCKET NUMBER: 9433
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 742-4343
; TELEFAX: (516) 742-4366
; TELETYPE: 230 901 SANS UR
; INFORMATION FOR SEQ ID NO: 50:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 496 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-313-075A-50

```

```

Query Match          10.4%; Score 270.5; DB 1; Length 496;
Best Local Similarity 23.2%; Pred. No. 1.1e-20;
Matches 117; Conservative 66; Mismatches 197; Indels 125; Gaps 16;

```

```

QY 33 VSSRDSCALPDPPTMGFPFEGTLOVLQRRKFLQMKRRKGYIKYKTHLFGRPYRVM 92
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 21 MSLRRRKSKRLPPGPPPLITGNHLGNHPPHSLAQKIHGPIKMLGQINTVVIS 80
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 93 GANNVRRL-----LGEHRLSVNH-PA-----SVRTLLGACLS--N 127
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 81 SSVAAEVLQKQDLTFSNRPVDPVYVHNRHSDFSVWLVPNSRWKTLRKIMNSIFSGNK 140
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 128 LHDSSHOKRKVIMQAFREALQCYVLIAEYVSCLEQMLSCGEGLLVYPEVKRLMFR 187
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 141 LDGNQHLRSKV-----QELIDYCK--CAKNEAV--DIGRAITFG 177
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 188 IMARILLGCEPGAGGDEQOLVEAFEEHTRNLPSLPIDVPFSGLYR----- 235
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 178 TTLNLT-----SNTIFSKDLTNPFSDSAKEREKELWNIMVE 213
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 236 -----GVKAR-----NLIHARIEENIRAKIRRLQATEPFGCG 267
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 214 AGKPNLVDPFPLEKIDPGIKRRMTNNPTKFGILSLID--RLKERLNR--DNAN 267
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 268 KDAQLLLIHSWGERLDMQALKOSSTELLFGGHETTASAAATSLITYGLVPHVLQVR 327
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 268 IDVALALNISOENPEIDIRNOIEQLCLDLFAAGTDTSTNLEWAMAEILLQNPHTLOKAQ 327
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 328 EIKSKGLCKSNQDNKLMETLEQKYGICVKEKTELRLNPVPVGGF--RYALKTFFELNGY 386
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 328 EELAQ--VIGKQVEAD---VGRIPYLRACIVKETLRHHPAAPLLIPKRVEDVELSTY 382
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 387 QIPKGMNVIYSICDTHVDADIFTNKEENPDREIVPHPEPDSR--FSFIFPGGGLRSCVCK 445
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 383 IIPKDSQVLYNVAIGNSDLDWENPLVFPERFWESEIDIRGDFLIPFGAORRIPCGL 442
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 446 EFAKILIKFTYELARHCWOLLNG 470
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 443 PLAMRRIPALGSLNSFNWKLKYG 467
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 13
US-09-292-768-68
; Sequence 68, Application US/09292768
; Patent No. 6194185
; GENERAL INFORMATION:
; APPLICANT: Croteau, Rodney B
; APPLICANT: Lupien, Shari L
; APPLICANT: Karp, Frank
; TITLE OF INVENTION: RECOMBINANT MATERIALS AND METHODS FOR THE PRODUCTION OF
; FILE REFERENCE: wsut13463
; CURRENT APPLICATION NUMBER: US/09/292,768

```

```

; CURRENT FILING DATE: 1999-04-14
; EARLIER APPLICATION NUMBER: 08/881,784
; EARLIER FILING DATE: 1997-06-24
; NUMBER OF SEQ ID NOS: 70
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 68
; LENGTH: 500
; TYPE: PRT
; ORGANISM: Artificial Sequence
; US-09-292-768-68

```

```

Query Match          10.0%; Score 262.5; DB 4; Length 500;
Best Local Similarity 21.9%; Pred. No. 8.4e-20;
Matches 115; Conservative 89; Mismatches 215; Indels 105; Gaps 19;

```

```

QY 7 LASALCTFVLPLLLFLAALKLMDLYCVSSRRSCALPLPPTMGFPFEGTLOWVLR-- 64
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 5 ISSAIIIVVITYITSLIIKQ-----KKPKQENLPFGPPKPLIGH-LHLLGKLP 56
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 65 RKLQMKRRKGYIKYKTHL-----FGRPYRVMGADN 96
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 57 QHALASVAKQYGPVAVHQLGVEFSVLSREATKEAMKLYDPACADRFESIGTKIMYDN 116
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 97 VRRILGEHRLSVHWPASVRTILGACLSLHDSHOKRKVIMQAFREALQCYLVI 156
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 117 -DIIIFSPY--SVHW-----ROMRKICVSELLSARNVAFGPIR 152
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 157 AEEVSSCLEQMLSCGEGLLVYPEVKRLMFRIA--MRILLCEPAGGDEQOLVAFE 215
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 153 QDEVSRLLGLHRSAAAG-----EAVDLTERIALTIGSIICRAFGSVIYDHETLV-- 203
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 216 EMTRNLSFLPIDVPFSGLYGVKARNLI-----HARIEENIRAKT--RLQATEP 263
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 204 ELVKDALSMASGELADMFSSSKLNLCLWNKSKLWMRRRVALLLEAIVEHKLKSGE 263
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 264 DGCKDALQLLIHSWGERLDM--QALKOSSTELLFGGHETTASAAATSLITYGLYPH 321
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 264 FGG-EDIIDVLFMOKDSQIKVPIITTKAIKAFIEDFISAGTESSTITLLWVMAELMRNPE 322
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 322 VLOKVEEIKSKGLCKSNQDNKLMETLEQKYGICVKEKTELRLNPVPVGGF--RYALKT 381
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 323 YMAKAQAEVRA--ALKGKD--WDVDVQVELKYMKSVMYKTRMHPPIPLIRSCRECG 377
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 382 ELNGYOIPKGMNVIYSICDTHVDADIFTNKEENPDREIVPHPEPDSR-----FSFIF 435
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 378 EVNGYIIPKARIMINIVWSGRNPLYWEKETWPERF-----DOYSRDPMGNDFFIIP 432
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 436 GGLRSCVCGKEFAKILIKFTYELARHCWOLLNG--PPTMKT 477
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 433 GAGRRRICPGINFGLANVEVPLAQLVHPDWKLAEGMNPSSMDMS 476
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 14
US-09-292-768-4
; Sequence 4, Application US/09292768
; Patent No. 6194185
; GENERAL INFORMATION:
; APPLICANT: Croteau, Rodney B
; APPLICANT: Lupien, Shari L
; APPLICANT: Karp, Frank
; TITLE OF INVENTION: RECOMBINANT MATERIALS AND METHODS FOR THE PRODUCTION OF
; FILE REFERENCE: wsut13463
; CURRENT APPLICATION NUMBER: US/09/292,768
; CURRENT FILING DATE: 1999-04-14
; EARLIER APPLICATION NUMBER: 08/881,784
; NUMBER OF SEQ ID NOS: 70
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 500
; TYPE: PRT

```

ORGANISM: Mentha piperita
us-09-292-768-4

Query Match 10.0%; Score 261.5; DB 4; Length 500;

Best Local Similarity 21.9%; Pred. No. 1.1e-19; Mismatches 216; Indels 105; Gaps 19;

Matches 115; Conservative 88; Mismatches 216; Indels 105; Gaps 19;

```
QY 7 LASALCTFVLPPLFLAALKLMIDLVCSSRDSCALPLPGTMGPFPPFGFTLOMVLOR-- 64
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 5 ISSAIIIVVTTTISILIIKQW-----RKPRQENLPQPPKPLIGH-LHLMGKLP 56
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 65 RKFLOMKRRKRYGFIYKTHL-----FGRPYVRMGADN 96
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 57 QHALASVAKQYGPVAHVQLEGEFVSVLSSREATKEAMKLVDPACADRESIGTKIMWYDN 116
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 97 VRRILGHRIVSVHMPASVRTILGAGCLSNLHDSHKORKVIMQAFSREALQCYLVI 156
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 117 -DDIIFSPY--SVHW-----RQMRKICVSELLSARNVRSFGFIR 152
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 157 AEEVSSCLEQWLSGCEGRLVYPEVKRLMERIA-MRILGCEPAGGDEQOLVEAFE 215
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 153 QDEVSRLLGHLSAAG-----EAVDLTERIATLTCISICRAFGSVIRDEHELY--- 203
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 216 EMTNRLFSPLIDVPFSGLYRGVAKANLI-----HARIEENIRAKI--RRLQATER 263
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 204 ELVKDALSMASGFEIADMFPSKILNLICWNKSKLMRMRRRVDAILEALVDEHKLKSGE 263
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 264 DGCKDALQLLIEHSWGERLDM--QALKOSTELLFGHETTASATSLITYLGLYPH 321
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 264 FGG-EDIIDVLFPMOKDSQIKVPTTNAIKAFIDTFSAGTETSTTTIWMALMRNPE 322
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 322 VLOKVEEIKSKGLCKSNODNKIDMETLEOLKYICVYIKETLRINPVPVGGFVALKTF 381
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 323 VMAKAAAEVRA---ALKGKTD--WDVDVQELKYMKSVMKETMRMHPPIPLIPRSCREEC 377
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 382 ELNGYQIPKGMNVYISICDTHVDADIFTNKEEFPNPRFIYPHEDASR-----FSFIPE 435
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 378 EVNGYTIIPNKRARIMINWMSGRNPLYMEKPEFTWPERF-----DOVSRDPMGNDFFIPE 432
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 436 GGGLRSCVCKEFAKILKFTVELARHCDWQLNG--PPTMKTS 477
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 433 GAGRRICPGLNPGLANVEVPLAQLLHYHFDWKLAEGMNPSMDMS 476
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 15
US-09-292-768-70
; Sequence 70, Application US/09292768
; Patent No. 6194185
; GENERAL INFORMATION:
; APPLICANT: Croteau, Rodney B
; APPLICANT: Croteau, Shari L
; APPLICANT: Karp, Frank
; TITLE OF INVENTION: RECOMBINANT MATERIALS AND METHODS FOR THE PRODUCTION OF
; FILE REFERENCE: W913463
; CURRENT APPLICATION NUMBER: US/09/292,768
; CURRENT FILING DATE: 1999-04-14
; EARLIER APPLICATION NUMBER: 08/881,784
; EARLIER FILING DATE: 1997-06-24
; NUMBER OF SEQ ID NOS: 70
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 70
; LENGTH: 500
; TYPE: PRT
; ORGANISM: Artificial Sequence
US-09-292-768-70
```

Query Match 10.0%; Score 261.5; DB 4; Length 500;
Best Local Similarity 21.9%; Pred. No. 1.1e-19;
Matches 115; Conservative 88; Mismatches 216; Indels 105; Gaps 19;

```
QY 7 LASALCTFVLPPLFLAALKLMIDLVCSSRDSCALPLPGTMGPFPPFGFTLOMVLOR-- 64
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 5 ISSAIIIVVTTTISILIIKQW-----RKPRQENLPQPPKPLIGH-LHLMGKLP 56
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 65 RKFLOMKRRKRYGFIYKTHL-----FGRPYVRMGADN 96
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 57 QHALASVAKQYGPVAHVQLEGEFVSVLSSREATKEAMKLVDPACADRESIGTKIMWYDN 116
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 97 VRRILGHRIVSVHMPASVRTILGAGCLSNLHDSHKORKVIMQAFSREALQCYLVI 156
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 117 -DDIIFSPY--SVHW-----RQMRKICVSELLSARNVRSFGFIR 152
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 157 AEEVSSCLEQWLSGCEGRLVYPEVKRLMERIA-MRILGCEPAGGDEQOLVEAFE 215
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 153 QDEVSRLLGHLSAAG-----EAVDLTERIATLTCISICRAFGSVIRDEHELY--- 203
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 216 EMTNRLFSPLIDVPFSGLYRGVAKANLI-----HARIEENIRAKI--RRLQATER 263
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 204 ELVKDALSMASGFEIADMFPSKILNLICWNKSKLMRMRRRVDAILEALVDEHKLKSGE 263
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 264 DGCKDALQLLIEHSWGERLDM--QALKOSTELLFGHETTASATSLITYLGLYPH 321
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 264 FGG-EDIIDVLFPMOKDSQIKVPTTNAIKAFIDTFSAGTETSTTTIWMALMRNPE 322
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 322 VLOKVEEIKSKGLCKSNODNKIDMETLEOLKYICVYIKETLRINPVPVGGFVALKTF 381
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 323 VMAKAAAEVRA---ALKGKTD--WDVDVQELKYMKSVMKETMRMHPPIPLIPRSCREEC 377
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 382 ELNGYQIPKGMNVYISICDTHVDADIFTNKEEFPNPRFIYPHEDASR-----FSFIPE 435
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 378 EVNGYTIIPNKRARIMINWMSGRNPLYMEKPEFTWPERF-----DOVSRDPMGNDFFIPE 432
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 436 GGGLRSCVCKEFAKILKFTVELARHCDWQLNG--PPTMKTS 477
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 433 GAGRRICPGLNPGLANVEVPLAQLLHYHFDWKLAEGMNPSMDMS 476
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

Search completed: November 6, 2001, 13:27:06
Job time: 9919 sec

cytochrome P450 CYP90 - Arabidopsis thaliana
 M:Contains: oxidoreductase (EC 1.-.-.-)
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
 C:Accession: S55379
 R:Sequences: M.; Nemeth, K.; Koncz, Z.; Nagy, F.; Koncz, C.
 submitted to the EMBL Data Library, May 1995
 A:Reference number: S55379
 A:Accession: S55379
 A:Molecule type: mRNA
 A:Residues: 1-472 <STP>
 A:Cross-references: EMBL:X87367; NID:g853718; PIDN:CAA60793.1; PID:g853719
 A:Gene: CYP90
 C:Genetics:
 C:Superfamily: Synechocystis cytochrome P450 slr0574; cytochrome P450 homology
 C:Keywords: chromoprotein; electron transfer; heme; iron; metalloprotein; monooxygenase;
 F:275-440/domain: cytochrome P450 homology <P45>
 F:418/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 17.2% Score 450; DB 1; Length 472;

Best Local Similarity 26.6%; Pred. No. 2.2e-27;
 Matches 134; Conservative 90; Mismatches 217; Indels 62; Gaps 15;

6 LLASALCFVPLPLFLAALKMDLYCVSSRDSCALPLPGTGMPPFGFTLQMV--- 61
 9 LLSIAGFFLL-----RRTRRRMGLPPGSLGLPLIGFTFOLIGATK 52
 QY 62 -LQRRFLQMRKRYGYLYKTHLGRPTVRVAGADNVRILLGHRVSVHWPASVETIL 120
 53 TENEPFIDEVARAGSGFMTHLFGEPITFSADETRNFVLQNGKLFESYPASICNL 112
 QY 121 GAGCSLNHDSHOKKVVITQASREALQCYLV-IAEVSCLQWLSGEGGLVLYP 179
 113 GKSHLLMKGSLHRMSLITMSFNSSITKDLMDLDRVRELDKSS---RVLM 168
 QY 180 EVKRLMFIARILLGEPGAGGEGDQVLEAFEMTNLFSLPIDVPFSLYR-GVK 238
 169 EAKITFELVYKQMSFDP-----GEMSESLRKYLLVIEGFFSLPL-FSTYRRAIQ 222
 QY 239 ARNLIAHEENIRAKIRLOATEPDGS--CKDALQILL--EHSMEGERLDMALQSS 294
 223 AR-----KVALLVVYVKKREBEERKEDKMDLAALADGFSDELYDF-----L 272
 QY 295 TELFGHETASATSLITLYGLVPHVLOKVRREIKSKGLCKSNODNKLMETLEOLK 354
 273 VALVAGETSTIMTLAVKFLTEPLALQLEE--HEKIRAKSDSYSLWSDYKSM 330
 QY 355 YIGVIRKTELNPVPGGERVALKTEPLNGYQIPKGMNVYSICDHVDADITNKEEF 414
 331 FTQCVNMTLLVANNIGVFRAMTDVEIKYKIPKGMKVSFRRAVHLDPNHFKDKRTF 390
 QY 415 NPDR-----IVHEDASRFSFPGGGLRSVCVKEPAKILKIFVELARHCDMOLING 470
 391 NPMRMQSNVYTGFSNV-----FTFPGCGPRICPGVELARVALSVFLRLVYGFSGW----- 441
 QY 471 PPMKTSPTVYVNLPRRTYF 493
 DB 442 VPADODKLVFFPTTRTORRYPF 464

RESULT 7
 T07859
 cytochrome P450 homolog - tomato

C:Species: Lycopersicon esculentum (tomato)
 C:Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 16-Feb-2001
 C:Accession: T07859

R:Bishop, G.J.; Harrison, K.; Jones, J.D.
 Plant Cell 8, 959-969, 1996

A:Title: The tomato Dwarf gene isolated by heterologous transposon tagging encodes the

A:Reference number: Z16181; MUID:96266705

A:Accession: T07859

A:Status: preliminary; translated from GB/EMBL/DBAJ

A:Molecule type: mRNA
 A:Residues: 1-464 <BITS>
 A:Cross-references: EMBL:U54770; NID:q1421740; PIDN:AA817070.1; PID:q1421741
 A:Experimental source: strain GCR738
 C:Genetics:
 A:Gene: dwarf
 C:Superfamily: Synechocystis cytochrome P450 slr0574; cytochrome P450 homology
 F:273-436/domain: cytochrome P450 homology <P45>

Query Match 16.9% Score 442; DB 2; Length 464;

Best Local Similarity 27.0%; Pred. No. 9e-27;
 Matches 125; Conservative 94; Mismatches 210; Indels 34; Gaps 10;

18 LLLFLAALKMDLYCVSSRDSCALPLPGTGMPPFGFTLQMVLRKRLQMRKRYGYF 77
 14 LCFICFALLRMNQVKNQKN-----LPPGTGMVPLFGTTEFLKSPFMKQBARAGS 67
 QY 78 IYKTHLFGRPVVRVAGADNVRILLGHRVSVHWPASVETILGAGCLSLHDSHOKK 137
 68 FFKSHLIGCPVYSMDSEINRYLVNEAKGLVPGYQSMIDILGKCNIAAVNGSAKTR 127
 QY 138 KVMQAFSEALQCYLV-IAEVSCLQWLSGEGGLVLYPEVKRLMFIARILLGC 196
 128 GALLSLISPTMTRDQLPRTIDFMRSHLFW--DNKVYIDQETKNKAFLSLKQIAGI 184
 QY 197 EPGPAGGEGDQVLEAFEMTNLFSLPIDVPFSLYGVAR---NLHARIEENIR 252
 185 ESTSLA-----GEMSEFNVLVGTSLPLNPNYHGFQARKIYNLLTLLEE--R 237
 QY 253 AKIRRLQATEPDGCGDALQILLHESMERGERLDMALQSSTELLFGHETASATSL 312
 238 RAKKEIQ-----HDMGLYLMNEATRFKLTDDMIDLIT-ILYSXETVSTSWMA 288
 QY 313 ITYGLVPHVLOKVRREIKSKGLCKSNODNKIDMETLEOLKTYGCVYKTRILMPVPG 372
 289 VKLHDPKVLLELRK--HMAIREKRPEDPIDVNDYRSMRFTRAVLLETSLATVNG 346
 QY 373 GFRVALKTEPLNGYQIPKGMNVYSICDHVDADITNKEENPDRFIVHEDASRFSF 432
 347 VLKRTQDMENIKYIIPKGMRIYVTRLENDPRILPDYSPNPMWMDKSLKHQ--SF 404
 QY 433 ITPGGGLRSCVGEKFAKILKIFVELARHCDMOLINGPPTK 475
 DB 405 LVFGGGRQCPKEKELGVAESTFLHYVTKYRMEIGDKLMK 447

RESULT 8
 H86185
 hypothetical protein [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
 C:Accession: H86185

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Cressy, T.H.; Dewar,

ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzia

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719

A:Accession: H86185

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-490 <STO>

A:Cross-references: GB:AE005172; NID:g2388581; PIDN:AA817462.1; GSPDB:GN00141

C:Genetics:
 A:Map position: 1
 C:Superfamily: Synechocystis cytochrome P450 slr0574; cytochrome P450 homology

QY 251 ---IRAKIRRLQATEPDPGGCKDALQLLIEHSWGERLDMQALKOSTELLFGHETTAS 307
 DB 257 YFARAKERR-----EASG--NDLTYLCQTEDDGNFSDADYNMHTFILMAAHDTSTS 310
 QY 308 AATSLITLYGLYPHVLQVBEIKSKGLCKSNODNKLDMETLEQLKYICVIRETLRLN 367
 DB 311 TATMAYOLAHPMQRORCRDESDRHG-----DGLDIESLEQLSDLVNMESTRIV 363
 QY 368 PPVGGGRVALKTFELNGXOIPKGMNVTISICDTHVDADIFTKKEENPRPFIYPHE-D 426
 DB 364 TPVOMAMQVATROTGLTLPKGTNVIAFGMNHRLPELTWDTLPFREFPRNEHK 423
 QY 427 ASRESFIFPGGKSCVCKEFAKI-----LKIFTVELARHCDWQLNGPPTMTSP 478
 DB 424 RHRAFAFPFGGVAKCIGMVFDDLEKITLIRLLRRYRLSLSR-DYQ-----PRMDYSA 477
 QY 479 TVPVDNLP 487
 DB 478 MPFMDQAP 486

RESULT 11

B84733
 C:Species: Arabidopsis thaliana (mouse-ear cross)
 C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
 C:Accession: B84733
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
 M.; Koo, H.; Moffat, K.S.; Cronlin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
 Euss, D.; Niekman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
 Nature 402, 761-768, 1999
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A:Reference number: A84420; MIMD:20083487
 A:Accession: B84733
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-489 <STO>
 A:Cross-references: GB:AE002093; NID:g3831452; PIDN:AAC69934.1; GSPDB:GN00139
 C:Genetics:
 A:Gene: At2g32440
 A:Map position: 2
 C:Superfamily: Synchocystis cytochrome P450 slr0574; cytochrome P450 homology

Query Match 15.5%; Score 406; DB 2; Length 489;
 Best Local Similarity 27.8%; Pred. No. 6.3e-24;
 Matches 128; Conservative 80; Mismatches 221; Indels 32; Gaps 12;

QY 44 LPPGTMGFPFGEITLQVY-----LQRRKFLQMKRRKG--FYKTHLFGRPVRYMGADN 96
 DB 44 LPPGDLMPVYIGNMWFLRAFKTSDPESFIQSYITRTGRTGTYKAHFGYPCVAVTTPET 103
 QY 97 VRILLGEHRLVSVHPASVVTLLGACLSNLHDSHKORKVYMAQFS-REALQCVLV 155
 DB 104 CRVLTDDDAF-HIGWPKSTMKILGRKSPVIGISPEEHRRLRLTSAPVNGPEALVYIOF 162
 QY 156 IAEVSSCLOMLSCGREGLLVYPEVKRLMRILLCGCEPPAGGDEQVYAEF 215
 DB 163 IEEVAVNDLEWMSKMEIEFLSH--LKRLLFKYIMYIFLSE-----SEIVMDSL 211
 QY 216 EMPTNL-----FSLPDIYVPSGLYGVAKRNLIHARIEENIRAKIRLQATEPDGCKDAL 271
 DB 212 REYNLNVGVAMGINLPFAVYHRLAKRKLVAA--QSIVTNRNRQKONISNRKDM 270
 QY 272 QLLIEHMERGERLDMQALKOSTELLEFGHETTASATSLITLYGLYPRVLOKVRREI 331
 DB 271 DNLIDVDMGRVLDDEILIDLLIMYLNAGHESGHTMTMATILMOHPMILQAKEE-- 328
 QY 332 SKGLCKSNODNKLDMETLEQLKYICVIRETLRLNPPVGGFRVALKTELANGYOIPK 391
 DB 329 QERIVKRRAPQKLTLEKREMYVLSQVIDETLKVITFSLAFREAKSDVQMDGYILPK 388
 QY 392 WNVYISICDTHVDADIFTKKEENPRPFIYPHEDASRFSFIPFGGGLRSCVCKEFAKI 451

DB 389 KAVLTFRRVNHLDPELYPPKFFDPSRMEGYTRKAG---TFLPFLGSHLCPGNDLALE 445
 QY 452 LKIFTVELARHCDWQLN-GPPTMTSPVYVDNLPAFT 491
 DB 446 ISIFLHHPILKYVERSNCGPVM--FLPHNRKDNCLARIT 485

RESULT 12

T02263
 C:Species: Zea mays (maize)
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
 C:Accession: T02263
 R:Winkler, R.G.; Helentjaris, T.
 Plant Cell 7, 1307-1317, 1995
 A:Title: The maize dwarf3 gene encodes a cytochrome P450-mediated early step in gibbe
 A:Reference number: Z14648; MIMD:96004534
 A:Accession: T02263
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-519 <MIN>
 A:Cross-references: EMBL:U32579; NID:g987266; PIDN:AAC49067.1; PID:g987267
 C:Experimental source: strain B73
 C:Genetics:
 A:Gene: dwarf3
 C:Function:
 A:Description: involved in an early step in gibberellin biosynthesis
 A:Pathway: gibberellin biosynthesis
 C:Superfamily: Synchocystis cytochrome P450 slr0574; cytochrome P450 homology
 F:325-488/Domain: cytochrome P450 homology <45>

Query Match 13.7%; Score 359; DB 1; Length 519;
 Best Local Similarity 26.3%; Pred. No. 3.2e-20;
 Matches 123; Conservative 89; Mismatches 213; Indels 42; Gaps 15;

QY 44 LPPGTMGFPFGEITLQVY-----FLQMKRRKG--FYKTHLFGRPVRYMGADN 96
 DB 72 LPPGEMGWPVYGVMAFLRAFKSGRPAFIASVRRFGRTGTVRFSMSSPVLVTTMAG 131
 QY 97 VRILLGEHRLVSVHPASVVTLLGACLSNLHDSHKORKVYMAQFS-ELQCVLV 155
 DB 132 CKQVLMDDDAFYT--GWPKATVALVGPSEFVAMPYDEHRRIRKLTAPINGFDALTYLPF 190
 QY 156 IAEVSSCLOMLSCGREGLLVY--PEVKRLMRILLCGCEPPAGGDE--QOLV 211
 DB 191 IDRTVYSSLRAMADHG--GSVEFLTELKRMPTKITVOLF-----GGADQATTRALE 240
 QY 212 EAFEMTRNLFSLPIDVPSGLYRG-VKARNLIHARIEENIRAKIRLQATEPDGCKDA 270
 DB 241 RSTYELNVMGRMAAINLP--GFAYRGALRLARRRLVA--VLQGYDERARRARAGVSGGYDM 298
 QY 271 LQLLIEHMERGERLDMQALKOSTELLEFGHETTASATSLITLYGLYPRVLOKVRREI 330
 DB 299 MRLTIAQDERGRHLDDEIDIVLYMYLNAGHESGHTMTMATVYLOINPMFARAKAQ 358
 QY 331 KSKGLCKSNODNKLDMETLEQLKYICVIRETLRLNPPVGGFRVALKTELANGYOIPK 390
 DB 359 EAIMSIPSSQ--RGLTLRFRKMEYLSQVIDETLRLVISEVSFROATRDVFNNGYLLPK 417
 QY 391 GNVVYISICDTHVDADIFTKKEENPRPFIYPHEDASRFSFIPFGGGLRSCVCKEFAKI 450
 DB 418 GKVQQLMYSVHMDQVYFPDPTKPSRMEGHSPPAG--TFLAFLGARLCPGNDLAKL 474
 QY 451 LKIFT-----VELARHCDWQLN-GPPTMTSPVYVDNLPAFT 491
 DB 475 EISVFLHHPILKYKLAR-----TNPCKRVYRLPBRPVDNCLAKIT 515

RESULT 13

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2001, 13:38:27 ; Search time 115.92 Seconds
(without alignments)
567.250 Million cell updates/sec

Title: US-09-668-482-32
Perfect score: 2612
Sequence: 1 MGLPALLASALCTFVLPILL.....FTYVPVNDLPARTFYFGDI 497

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: SP_TREMBL_16:*
2: sp_archaea:*
3: sp_bacteria:*
4: sp_fungi:*
5: sp_human:*
6: sp_invertebrate:*
7: sp_mammal:*
8: sp_mhc:*
9: sp_organelle:*
10: sp_phase:*
11: sp_plant:*
12: sp_rodent:*
13: sp_unclassified:*
14: sp_vertebrate:*
15: sp_virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2609	99.9	497	11	09R1F4	09R1F4 mus musculus
2	2057.5	78.8	492	13	09P1B4	09P1B4 gallus galli
3	1758.5	67.3	492	13	09J323	09J323 xenopus lae
4	1048.5	40.1	512	4	09NR63	09NR63 homo sapien
5	1036	39.7	525	4	09NP41	09NP41 homo sapien
6	668.5	25.6	444	2	059990	059990 synechocyst
7	569	21.8	477	10	09LYV7	09LYV7 arabidopsis
8	538	20.6	107	13	09P1G2	09P1G2 gallus galli
9	517.5	19.8	463	10	09F176	09F176 arabidopsis
10	511.5	19.6	485	10	09SJH2	09SJH2 arabidopsis
11	501.5	19.2	482	10	081077	081077 arabidopsis
12	494.5	18.9	457	10	065624	065624 arabidopsis
13	461	17.6	513	10	064989	064989 arabidopsis
14	461	17.6	513	10	09SC09	09SC09 arabidopsis
15	455.5	17.4	465	10	09FMA5	09FMA5 arabidopsis
16	432	16.5	443	10	09LHK2	09LHK2 arabidopsis
17	418.5	16.0	465	10	09LH81	09LH81 arabidopsis
18	418	16.0	457	10	09M066	09M066 arabidopsis
19	418	16.0	474	10	09LKH7	09LKH7 phaseolus a

20	418	16.0	524	10	023242	023242 arabidopsis
21	407.5	15.6	464	10	09LIC5	09LIC5 arabidopsis
22	406	15.5	489	10	09ZV72	09ZV72 arabidopsis
23	404	15.5	478	10	09LNT3	09LNT3 arabidopsis
24	400.5	15.3	496	10	09FQY4	09FQY4 cucurbita m
25	380	14.5	475	2	09X7G9	09X7G9 myxococcus
26	374	14.3	504	10	09SNC3	09SNC3 oryza sativ
27	358.5	13.7	511	10	09SMT1	09SMT1 liqidambar
28	351	13.4	486	11	09J1Y3	09J1Y3 mus musculu
29	351	13.4	503	11	09J1P8	09J1P8 mus musculu
30	350.5	13.4	513	10	09FVB8	09FVB8 brassica na
31	342	13.1	520	10	09FVC0	09FVC0 brassica na
32	341	13.1	520	10	09FVB9	09FVB9 brassica na
33	340	13.0	318	10	09LYV3	09LYV3 arabidopsis
34	336.5	12.9	735	10	09LGI7	09LGI7 arabidopsis
35	336	12.9	521	10	09XFW2	09XFW2 lycopersico
36	333	12.7	518	10	09F138	09F138 arabidopsis
37	329.5	12.6	512	10	09FX29	09FX29 arabidopsis
38	327	12.5	498	5	016806	016806 drosophila
39	318.5	12.2	474	10	09SHY7	09SHY7 arabidopsis
40	318.5	12.2	491	5	09NGK3	09NGK3 tribolium c
41	318.5	12.2	513	10	09SBP8	09SBP8 populus trl
42	318	12.2	497	5	076478	076478 diptoptera
43	317.5	12.2	459	2	059163	059163 anabaena va
44	315.5	12.1	495	10	09TOK2	09TOK2 arabidopsis
45	314	12.0	464	10	004949	004949 arabidopsis

ALIGNMENTS

RESULT	ID	PRELIMINARY:	PRT:	497 AA.
09R1F4	09R1F4			
AC	09R1F4			
DT	01-MAY-2000 (TREMBLrel. 13, Created)			
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)			
DT	01-MAR-2001 (TREMBLrel. 16, Last annotation update)			
DE	CTOCHROME P450 RETINOIC ACID METABOLIZING ENZYME P450RA.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.			
OX	NCBI_Taxid=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=TOOTH;			
RX	Pubmed=11063033;			
RA	Paine C.T., Paine M.L., Sneed M.L.;			
RT	"Identification of tufelin- and amelogenin-interacting proteins using the yeast two-hybrid system."			
RL	Connect. Tissue Res. 38:257-267(1998).			
CC	-1- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) = ROH + OXIDIZED FLAVOPROTEIN + H(2)O (BY SIMILARITY).			
CC	-1- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM (BY SIMILARITY).			
CC	-1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.			
CC	EMBL: AF115769; AD17217.1; ..			
DR	InterPro:IPR001128; ..			
DR	Pfam: PF00067; P450; 1.			
DR	PRINTS: PR00385; P450.			
DR	PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.			
KW	Electron transport; Endoplasmic reticulum; Heme; Membrane; Microsome; Monooxygenase; Oxidoreductase.			
SQ	SEQUENCE 497 AA; 56191 MW; 67CB167A997842C1 CRC64;			
Qy	1 MGLPALLASALCTFVLPILLFLALAKIMLDYCVSSRRSCALPLPPTMGPPFGETIOM 60	99.9%; Score 2609; DB 11; Length 497;		
Db	1 MGLPALLASALCTFVLPILLFLALAKIMLDYCVSSRRSCALPLPPTMGPPFGETIOM 60	Best Local Similarity 99.8%; Pred. No. 1.3e-205; Matches 496; Conservative 1; Mismatches 0; Indels 0; Gaps 0;		

```

QY 61 VLORRKFLOMRKRRKGYFTKTHLFGPTVRVGMADNVRILLGHEHRLYSVHMPASVRTIL 120
D6 61 VLORRKFLOMRKRRKGYFTKTHLFGPTVRVGMADNVRILLGHEHRLYSVHMPASVRTIL 120
QY 121 GAGCLSNLHDSHROKRVKVMQAFSREALQCYVLVIAEVSSCLEQWLSCGERGLLYPE 180
D6 121 GAGCLSNLHDSHROKRVKVMQAFSREALQCYVLVIAEVSSCLEQWLSCGERGLLYPE 180
QY 181 VKRLMFRIAMILLGCEPGAGGEDQOLVEAFEEEMTRNLFSLPIDVPFSGLYRGVAKR 240
D6 181 VKRLMFRIAMILLGCEPGAGGEDQOLVEAFEEEMTRNLFSLPIDVPFSGLYRGVAKR 240
QY 241 NLHARIEENIRAKIRRLQATEPDGCKDALQILLIHSWEMGERLDMQALKOSTTELLFG 300
D6 241 NLHARIEENIRAKIRRLQATEPDGCKDALQILLIHSWEMGERLDMQALKOSTTELLFG 300
QY 301 GHETTASAASTLYTLGLYPRVLOKVRBEIKSKGLCKSNODNKLDMETTELQKYICGVI 360
D6 301 GHETTASAASTLYTLGLYPRVLOKVRBEIKSKGLCKSNODNKLDMETTELQKYICGVI 360
QY 361 KETRLNLPVPGGFRVALKTFELNGYQIIPKGMNVIYSICDTHVADIFTNKEEENPDREI 420
D6 361 KETRLNLPVPGGFRVALKTFELNGYQIIPKGMNVIYSICDTHVADIFTNKEEENPDREI 420
QY 421 VPHBEDASRSFIFPGGGLSCVCKEFAKILKIFVELARHCDMQLNGPPTKTSPTV 480
D6 421 VPHBEDASRSFIFPGGGLSCVCKEFAKILKIFVELARHCDMQLNGPPTKTSPTV 480
QY 481 YPVNDLPARTFYOGDI 497
D6 481 YPVNDLPARTFYOGDI 497

RESULT 2
ID 09PUB4 PRELIMINARY; PRT; 492 AA.
AC 09PUB4:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, last sequence update)
DE CYTOCHROME P450 26 (EC 1.14.-.-) (RETINOIC ACID-METABOLIZING
DE CYTOCHROME) (P450RAI) (RETINOIC ACID 4-HYDROXYLASE).
GN CYP26A1.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauromorpha; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RA Swindell E.C., Thaller C., Sockanathan S., Petkovich M., Jessell T.M.,
RA Eichele G.;
RT "Complementary Domains of Retinoic Acid Production and Degradation in
RT the Early Chick Embryo.";
RL Dev. Biol. 0-0-0(2000).
CC -1- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) = ROH +
CC OXIDIZED FLAVOPROTEIN + H(2)O (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC EMBL: AF199462; AAF09250.1; -.
DR InterPro: IPR001128; -.
DR Pfam: PF00067; P450; 1.
DR PRINTS: PR00385; P450.
DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
KW Electon transport; Endoplasmic reticulum; Heme; Membrane; Microsome;
KW Monooxygenase; Oxidoreductase.
SQ SEQUENCE 492 AA; 55264 MW; 7P28B72E75C232FB CRC64;

```

Query Match 78.8%; Score 2057.5; DB 13; Length 492;
Best Local Similarity 80.1%; Pred. No. 2.1e-160;

```

Matches 398; Conservative 40; Mismatches 54; Indels 5; Gaps 3;
QY 1 MGLPALLASALCTEVLPLLLFLAALKMDLYCVSSRDSALPLPPTMGPFPGTETL 60
D6 1 MGLPALLASALCTEVLPLLLFLAALKMDLYCVSSRDSALPLPPTMGPFPGTETL 60
QY 61 VLORRKFLOMRKRRKGYFTKTHLFGPTVRVGMADNVRILLGHEHRLYSVHMPASVRTIL 120
D6 61 VLORRKFLOMRKRRKGYFTKTHLFGPTVRVGMADNVRILLGHEHRLYSVHMPASVRTIL 120
QY 121 GAGCLSNLHDSHROKRVKVMQAFSREALQCYVLVIAEVSSCLEQWLSCGERGLLYPE 180
D6 121 GAGCLSNLHDSHROKRVKVMQAFSREALQCYVLVIAEVSSCLEQWLSCGERGLLYPE 180
QY 181 VKRLMFRIAMILLGCEPGAGGEDQOLVEAFEEEMTRNLFSLPIDVPFSGLYRGVAKR 240
D6 181 VKRLMFRIAMILLGCEPGAGGEDQOLVEAFEEEMTRNLFSLPIDVPFSGLYRGVAKR 240
QY 241 NLHARIEENIRAKIRRLQATEPDGCKDALQILLIHSWEMGERLDMQALKOSTTELLFG 300
D6 241 NLHARIEENIRAKIRRLQATEPDGCKDALQILLIHSWEMGERLDMQALKOSTTELLFG 300
QY 301 GHETTASAASTLYTLGLYPRVLOKVRBEIKSKGLCKSNODNKLDMETTELQKYICGVI 360
D6 301 GHETTASAASTLYTLGLYPRVLOKVRBEIKSKGLCKSNODNKLDMETTELQKYICGVI 360
QY 361 KETRLNLPVPGGFRVALKTFELNGYQIIPKGMNVIYSICDTHVADIFTNKEEENPDREI 420
D6 361 KETRLNLPVPGGFRVALKTFELNGYQIIPKGMNVIYSICDTHVADIFTNKEEENPDREI 420
QY 421 VPHBEDASRSFIFPGGGLSCVCKEFAKILKIFVELARHCDMQLNGPPTKTSPTV 480
D6 421 VPHBEDASRSFIFPGGGLSCVCKEFAKILKIFVELARHCDMQLNGPPTKTSPTV 480
QY 481 YPVNDLPARTFYOGDI 497
D6 481 YPVNDLPARTFYOGDI 497

RESULT 3
ID 093323 PRELIMINARY; PRT; 492 AA.
AC 093323:
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, last sequence update)
DE CYTOCHROME P450 26 (EC 1.14.-.-) (RETINOIC ACID-METABOLIZING
DE CYTOCHROME) (P450RAI) (RETINOIC ACID 4-HYDROXYLASE).
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae;
OC Xenopodidae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RA Chen J.L., Grunz H., Panitz F., Pieler T., Hollmann T.;
RA Submitted (APR-1998) to the EMBL/Genbank/DBJ databases.
RL -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC EMBL: AF057566; AAC25158.1; -.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC InterPro: IPR001128; -.
CC Pfam: PF00067; P450; 1.
CC PRINTS: PR00385; P450.
CC PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
KW Heme; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 492 AA; 55459 MW; D14BB7651BF2D3E CRC64;

```

Query Match 67.3%; Score 1758.5; DB 13; Length 492;
Best Local Similarity 68.1%; Pred. No. 6.9e-136;
Matches 339; Conservative 63; Mismatches 89; Indels 7; Gaps 4;


```
Db 1 MDLYTLTSLAFCYALPLLLLLTAATLMEVYCLRRKDAACANLPPTGMLPFFGCTLOK 60
QY VLORRKFLQKRRKRYGFIYKTHLFGRPYVMGADVNRILLGHRVLSVHPASVYRLL 120
Db 61 VLORRRFLQKRRKRYGFIYKTHLFGSPYAVTGAENVROILGMEHHLVSVHPASVYRLL 120
QY 121 GAGCLSLNHDSSHOKRKVIMQAFSRBALQCYLVLAEEVSSCLEOMLSCGEGNLLYYPE 180
Db 121 GAGCLSLNHDSSHOKRKVIMQAFSRBALQCYLVLAEEVSSCLEOMLSCGEGNLLYYPE 179
QY 181 VKLMPFIARIRLLIGCEPAGGAGEBOOLVEAFEEETRNLFLSLPIDVPSGLYRGVKAR 240
Db 180 IKRMFEIARLLKQGP-ORMDRODEETLEAFEBKSRNLFLSLPIDVPSGLYRGVKAR 238
QY 241 NLHARIEENIRAKIRRLQATEPDDGGCKDALQLLIEHSWGERLDMQALKOSTELLFG 300
Db 239 NLHARIEENIRAKIRRLQATEPDDGGCKDALQLLIEHSWGERLDMQALKOSTELLFG 294
QY 301 GHETTASATSLTYLGLYVLAQVREIKSKGL-CKSNQDNKLDMEETLEQKTYIGV 359
Db 295 GHETTASATSLTYLGLYVLAQVREIKSKGL-CKSNQDNKLDMEETLEQKTYIGV 354
QY 360 IKETRLNPPVPGGFYALKTFFELNGQIPKGNVYISICDTHDVAADIFTNKEEFNDRF 419
Db 355 IKETRLNPPVPGGFYALKTFFELNGQIPKGNVYISICDTHDVAADIFTNKEEFNDRF 414
QY 420 IVPHPEDASRFSFIPFGGGRSCVCKEFAKILKIFTELARHCMDQMLNGPPTMKTSP 479
Db 415 LTPLEPDRSSRGFIPFGGGRSCVCKEFAKILKIFTELARHCMDQMLNGPPTMKTSP 474
QY 480 VYPVNDLPARTYFQDGI 497
Db 475 ICPVNDLPARTYFQDGI 492

RESULT 4
Q9NR63 PRELIMINARY; PRT; 512 AA.
AC Q9NR63;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE CYTOCHROME P450 RETINOID METABOLIZING PROTEIN P450RA1-2.
GN CYP26B1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20300913; PubMed=10823918;
RA White J.A., Ramshaw H., Taimi M., Stangle W., Zhang A., Everingham S.,
RA Creighton S., Tam S.-P., Jones G., Petkovitch M.;
RT "Identification of the human cytochrome P450, P450RA1-2, which is
RT predominantly expressed in the adult cerebellum and is responsible for
RT all trans-retinoic acid metabolism.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:6403-6408(2000).
CC -1- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) = ROH +
CC OXIDIZED FLAVOPROTEIN + H(2)O (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC EMBL, AF252297; AAF76003.1; -.
DR InterPro: IPR001128; -.
DR Pfam: PF00067; P450; 2.
DR PRINTS: PR00385; P450.
DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
KW Election transport; Endoplasmic reticulum; Heme; Membrane; Microsome;
KW Monooxygenase; Oxidoreductase.
SQ SEQUENCE 512 Aa: 57512 MW: A06DD19944E6726F CXC64;
```

Query Match 40.1%; Score 1048.5; DB 4; Length 512;

```
Best Local Similarity 44.4%; Pred. No. 1,2e-77;
Matches 220; Conservative 88; Mismatches 164; Indels 23; Gaps 9;

QY 7 LASALCTEVLPL----LLEALAKIMLDLYCVSSRDSCALPLPPTMGFFPFGCTLOAVL 62
Db 8 LVSALATLAACLVSTVLLIAVSQQLMQLRMATRDQSKPLPKSGMFGPLIGGHWLL 67
QY 63 QRRKFLQKRRKRYGFIYKTHLFGRPYVMGADVNRILLGHRVLSVHPASVYRLL 122
Db 68 QGSGFQSSRRERYGVNFKTHLGRPLIRYGAENVKILMGHHLVSTEWRSRMLLGP 127
QY 123 GCLSLNHDSSHOKRKVIMQAFSRBALQCYLVLAEEVSSCLEOMLSCGEGNLLYYPEVK 182
Db 128 NVVNSNIGDIHNKRRVSKIFPSHEALSYLPKQIVLQIDTLRAWSHPE-AINYYQBAQ 186
QY 183 RLMFRIARIRLLIGCE-PGAGGAGEBOOLVEAFEEETRNLFLSLPIDVPSGLYRGVKAR 241
Db 187 KLTFRMAIRVLLIGFSIP----EEDGLFEVYQFVNDVFSLPVLDLPSPGYRRGQIAQ 241
QY 242 LNHARIEENIRAKIRRLQATEPDDGGCKDALQLLIEHSWGERLDMQALKOSTELLFG 301
Db 242 LLOKGLEKAIREFKQCTQCKD-----YLDMLDILSSKHEKHEMTMOELKQDTLEIFAA 297
QY 302 HETTASATSLTYLGLYVLAQVREIKSKGL-CKSNQDNKLDMEETLEQKTYIG 357
Db 298 YATTASASLIMQLKHPTLEKLDRELRAHGLHSGGCPG-EGTIRLDYLSGLRYLD 355
QY 358 CVIKETRLNPPVPGGFYALKTFFELNGQIPKGNVYISICDTHDVAADIFTNKEEFNDR 417
Db 356 CVIKEVRLFTPIPSGGYRVYLAQVREIKSKGL-CKSNQDNKLDMEETLEQKTYIG 415
QY 418 RFTVHPEDA-SRFSFIPFGGGRSCVCKEFAKILKIFTELARHCMDQMLNGP-PTMK 475
Db 416 RFSQARSEDKDGRFHYLPFGGGRSCVCKEFAKILKIFTELARHCMDQMLNGP-PTMK 475
QY 476 TSPYVYVNDLPARF 490
Db 476 LVPLVHPVDGLSYKF 490

RESULT 5
Q9NP41 PRELIMINARY; PRT; 525 AA.
AC Q9NP41;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE WOGSC:H_NH0493L16.1 PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99063792; PubMed=9847074;
RA Sulston J.E., Waterston R.;
RT "The sequence of Homo sapiens BAC clone RP11-493L16.";
RT "Toward a complete human genome sequence.";
RL Genome Res. 8:1097-1108(1998).
CC [2]
CC SEQUENCE FROM N.A.
CC Bourne S., Bauer C., Page K., Jones T.;
CC "The sequence of Homo sapiens BAC clone RP11-493L16.";
CC Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Waterston R.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) = ROH +
CC OXIDIZED FLAVOPROTEIN + H(2)O (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC EMBL, AC007002; AAF65576.1; -.
```

DR InterPro: IPR001128; -
 DR Pfam: PF00067; P450; 1.
 DR PRINTS: PR00385; P450.
 DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
 DR Electrophoretic transport, Endoplasmic reticulum; Heme; Membrane; Microsome;
 KW Monooxygenase; Oxidoreductase.
 SQ SEQUENCE 525 AA; 59124 MW; 71D47B6752A60315 CRC64;

Query Match 39.7%; Score 1036; DB 4; Length 525;
 Best Local Similarity 43.3%; Pred. No. 1.3e-76;
 Matches 220; Conservative 89; Mismatches 163; Indels 36; Gaps 10;

QY 7 LASALCTFVLP-LLFLAALKMDLYCVSSRDSRSCALPLPGTMGFPFGTGLAVL 62
 DB 8 LVSAALATLACLVSVTLIAVSSQQLMDLRNATRDCKSLPIPGSMGFPILIGTGHLL 67
 QY 63 QR-----RRFLQMKRRKRGFYIKTHLFGRPYAVMGADVNRILLGEHRLVS 109
 DB 68 QKCTLRTRVWLPQSGFSSRRKRYGNVEKTHLGRPLRVTAENVRKILMEHHLVS 127
 QY 110 VHPASVITLIGAGCLSNLHDSHKORKVIMQAFSEALQCYLVLAEEVSSGLEMLS 169
 DB 128 TEMPRSTRMLGPRVTSNSIGDIHRNKRKYSKTFSEHALESYLPKIQLYQDTLRAMSS 187
 QY 170 CGERGLLYEVKRMFRIMRILGCE-DGAPGGDEQOLVEAFEMTNLPLPIDV 228
 DB 188 HPE-ALINVYOEAQKLFPRMAIRVLGFSIP-----EEDLGHLEFVYQOFVNDVSLPYDL 241
 QY 229 PFGSLYGVARNLHARIEINIRAKIRLQATNEPDGCKDALQILLHESNEBERLDMQ 288
 DB 242 PFGSGRIGIARQILQKGLKAIKREKLCQCKD---YLDLDELLESSKEHEKEMTQ 297
 QY 289 ALKOSTELLFGHETTASATSLITYGLYPHYLOKVEEIKSGKL---CKSNODNK 344
 DB 298 ELKDGLTELEFAVAYTASASTSLIMQLKHPTVLEKRLBELNAHGLHSGGCC-EGT 355
 QY 345 LQMETLEOLKTYICVKEITRLNRPVGGFRVALKTFELNGYQIPKMWYISICTHDV 404
 DB 356 LKIDTISGLYLDVCVKEVRLFTPLISGRTYVQLQTFELDGFQIPKMSWYISIRTDHT 415
 QY 405 ADLTNKEEFNDRFIVPHPEDA-SRFSFIPFGGLRSCVGEKFAKLLIKFYTELARHC 463
 DB 416 APYFKOVANVEDPDRFGQARSDEKDGRRHYIPFGGVRTCGLKHLAKFLVLAELASTS 475
 QY 464 DMQLNGP-PTMKTSPYVPVNLPAF 490
 DB 476 RELATRTPRITLVPVLPVDELGVKF 503

RESULT 6
 ID 059990 PRELIMINARY; PRT; 444 AA.
 AC 059990;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, last annotation update)
 DE PUTATIVE CYTOCHROME P450 120.
 GN CYP120 OR CYP OR SLR0574.
 OS Synechocystis sp. (strain PCC 6803).
 OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
 OX NCBI_TaxID=1148;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PCC6803;
 RL Tabata S.;
 RA Submitted (Aug-1995) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96127529; PubMed=8590279;
 RA Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,
 RA Sugitara M., Tabata S.;
 RT *Sequence analysis of the genome of the unicellular cyanobacterium

RT Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
 RT region from map positions 648 to 928 of the genome.";
 RL DNA Res. 2:153-166(1995).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97061201; PubMed=8905231;
 RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
 RA Miyajima N., Hirose M., Sugitara M., Nakazaki N., Nishizaki S.,
 RA Hosouchi T., Matsuo A., Muraki A., Nakazaki N., Nishizaki S.,
 RA Shimo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
 RA Tabata S.;
 RT *Sequence analysis of the genome of the unicellular cyanobacterium
 RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
 RT entire genome and assignment of potential protein-coding regions.";
 RL DNA Res. 3:109-136(1996).
 CC -1. STABILITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 DR EMBL: D64003; BAA10496.1; -
 DR InterPro: IPR001128; -
 DR Pfam: PF00067; P450; 1.
 DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
 KW Hypothetical protein; Heme; Monooxygenase; Oxidoreductase.
 SQ SEQUENCE 444 AA; 50578 MW; 8F62A9ED3B54BDC CRC64;

Query Match 25.6%; Score 668.5; DB 2; Length 444;
 Best Local Similarity 34.1%; Pred. No. 1.3e-46;
 Matches 154; Conservative 87; Mismatches 181; Indels 29; Gaps 8;

QY 41 ALPLPCTMGFPFGTGLQVLYORRFLQMKRRKRYGYIKTHLFGRPYRVMGADVNRRI 100
 DB 10 SLPIPDGFLPWTGETLNF-LNDGDFGKKRQOFGPIFKTRLEGKWNIFISGALNARFL 68
 QY 101 LLEGHRLVSVHPASVITLIGAGCLSNLHDSHKORKVIMQAFSEALQCYLVLAEEV 160
 DB 69 FTKQETFOATWPLSTILGLPNALATGEBIHSRKILYQAFPLPTLDSYLPKMGIV 128
 QY 161 SSCLEQMLSCGERGLLYPEVKRLMFRIMRILGCEPRGAGGDEQOLVEAFEEETRN 220
 DB 129 QGYLEQWKANE-VIVYQPLRRYTFVNAATLFMGKAV-----SQNPQLPMEETIYG 180
 QY 221 LFSLPIDVPSGLYRGVKANLHARIEINIRAKIRLQATNEPDGCKDALQILLHESWE 280
 DB 181 LFSLPIDVPSGLYRGVKANLHARIEINIRAKIRLQATNEPDGCKDALQILLHESWE 280
 QY 281 RGERLDMQALKOSTELLFGHETTASATSLITYGLYPHYLOKVEEIKSGGLCKSN 340
 DB 235 NNQPLSPLEKQDILLFLFGHETITLSALSPCLLGOHDIKERVQEQN-----KLQ 288
 QY 341 QDNKLMETLEOLKTYICVKEITRLNRPVGGFRVALKTFELNGYQIPKMWYISICTD 400
 DB 289 LSOELTAETLKKMPYLDQVLAQVLRILRPVGGFRRELIDQCQFGFHPKGLVSYIQS 348
 QY 401 THDVADIFTNKEEFNDRFIVPHPEDA-SRFSFIPFGGLRSCVGEKFAKLLIKFYTEL 456
 DB 349 THADPDVIRPPEKFDERT--PDGSATHNPPFAVPPGGLREGLGKFAHLEMKLFA 405
 QY 457 VELARHCQMLNGP-PTMKTSPYVPVNL 486
 DB 406 TRLIQGFQDWTLLPGQNLVELVPSRPKNL 436

RESULT 7
 ID 09LVY7 PRELIMINARY; PRT; 477 AA.
 AC 09LVY7;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, last annotation update)
 DE CYTOCHROME P450-LIKE.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 OC Brassicales; Brassicaceae; Arabidopsis.

```

OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-COLUMBIA:
RX MEDLINE=20181125; PubMed=10718197;
RA Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Kotani H.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence
RT features of the regions of 3,076,755 bp covered by sixty P1 and TAC
RT clones";
RL DNA Res. 7:31-63(2000).
CC -1- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) = ROH +
CC OXIDIZED FLAVOPROTEIN + H(2)O (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL: AB018112; BAA96885.1; -.
DR InterPro: IPR001128; -.
DR Pfam: PF00067; P450; 1.
DR PRINTS: PR00385; P450.
DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
KW Electron transport; Endoplasmic reticulum; Heme; Membrane; Microsome;
KW Monooxygenase; Oxidoreductase.
SQ SEQUENCE 477 AA; 54850 MW; 304B4B0C4970E405 CRC64;

Query Match 21.8%; Score 569; DB 10; Length 477;
Best Local Similarity 31.6%; Pred. No. 2.1e-38;
Matches 154; Conservative 82; Mismatches 215; Indels 36; Gaps 14;

OY 18 LILFLALAKLMDLYCVSSRRSCALP-LPPGTGFPFGFTLQMVLDOR-----KFLQMK 71
DB 7 LILFLSSILSLILKHLKSHSYPLNPPGNTGLPLIGSFSLSRGCRGHEPKFTTDR 66
OY 72 RRYK-----GFIYKTHLEGRPTVRYMGADNVRLLIGHRILVSVHPASVETILGAGCLN 127
DB 67 VRRSSSSSCVCFKHLGSPYAVVTGASGNKFLFTNENKLVSWPDSVYKIFP-----SS 122
OY 128 LHDSSHKORKKVM---QATSRRLQCVLYIAEVSQCLE-QWLSGCEKGLVYPEVKR 183
DB 123 MOTSSKEARLRILMISOEMKPEALRRYGVMDIQRHFETEW--ANODQVIVFPLTK 180
OY 184 LMFRLAMRIILGCEGPGAGGEDQOLVEAFEMTRNLSPIDVPFSGLYRGVAKANLI 243
DB 181 FTFSIACSPISME-DPA-----RYROLEQDPNTYAVGIFSPIDLPGRFRMAIKASRL 235
OY 244 HARLEENIRAKIRLOATEPDGCKDALOLLIEHSWGERLDMQALQOSTELLFGGHE 303
DB 236 RKEVASAIYRQKKEELKA-----GKALEHNDILSHMLMNGETKD-EDLADKILIGLIGHD 290
OY 304 TTASATSLITYLGLYVLOKVRBEIKSKULLCSNODNKLDMETLEOLKIYIGCVIKET 363
DB 291 TASIYCTFVVNVLAEPHYVORVLOE--QKELIKKKKEKELRMEDIEKMYSMNVACEV 348
OY 364 IRLNPVGVGGFRVALKTELNGYOIPKGMNVTISCDTHVADITFNKEENPDRFIYPH 423
DB 349 MIVPLPLSGTFRFEDALDHSFGKFTIPKGMKLYWSATATHAMPDTPPEPEREPNPFBSG 408
OY 424 PEDASRFSPFGGRLSCVGNKFAKILKIFTVELARHCWQ--LLNCPPTMKTSPTVYP 482
DB 409 PEP---YTYVPGGGPRMPCGKEVARLELLIFMNLVNRKWEKVPENKNIIVDPLIP 465
OY 483 VDNLPAR 489
DB 466 DKGLPIR 472

RESULT 8
ID 09PUG2 PRELIMINARY; PRT; 107 AA.
AC 09PUG2;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)

```

```

DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE CYP26 (FRAGMENT).
GN CYP26.
OS Gallus gallus (Chicken).
OC Gallus gallus; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Eukaryota; Metazoa; Neochordata; Galliformes; Phasianidae; Phasianinae;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RA Martinez-Ceballos E., Burdall C.A.;
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) = ROH +
CC OXIDIZED FLAVOPROTEIN + H(2)O (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL: AF185266; AAD56546.1; -.
DR HSSP: P14779; 1B07.
DR InterPro: IPR001128; -.
DR Pfam: PF00067; P450; 1.
DR PRINTS: PR00385; P450.
DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
KW Electron transport; Endoplasmic reticulum; Heme; Membrane; Microsome;
KW Monooxygenase; Oxidoreductase.
FT NON_TER 1
FT TER 107
SQ SEQUENCE 107 AA; 12009 MW; CA68BF0D598BC474 CRC64;

```

```

Query Match 20.6%; Score 538; DB 13; Length 107;
Best Local Similarity 90.7%; Pred. No. 8.6e-37;
Matches 97; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

```

```

OY 373 GFRVALKTEPLNCOYIPKGMNVTISCDTHVADITFNKEENPDRFIYPHEDASRFSE 432
DB 1 GFRVALKTEPLNCOYIPKGMNVTISCDTHVADITFNKEENPDRFIYPHEDASRFSE 60
OY 433 IPFGGRLSCVGNKFAKILKIFTVELARHCWQ--LLNCPPTMKTSPT 479
DB 61 IPFGGRLSCVGNKFAKILKIFTVELARHCWQ--LLNCPPTMKTSPT 107

```

```

RESULT 9
ID 09FH76 PRELIMINARY; PRT; 463 AA.
AC 09FH76;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE CYTOCHROME P450.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-COLUMBIA:
RX MEDLINE=20181125; PubMed=10718197;
RA Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Kotani H.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence
RT features of the regions of 3,076,755 bp covered by sixty P1 and TAC
RT clones";
RL DNA Res. 7:31-63(2000).
DR EMBL: AB020744; BAB10255.1; -.
SQ SEQUENCE 463 AA; 52366 MW; CCD17293F553FB12 CRC64;

```

```

Query Match 19.8%; Score 517.5; DB 10; Length 463;
Best Local Similarity 29.3%; Pred. No. 3.4e-34;
Matches 143; Conservative 88; Mismatches 212; Indels 45; Gaps 11;

```


Query Match 19.2%; Score 501.5; DB 10; Length 482;
 Best Local Similarity 28.2%; Pred. No. 7,4e-33;
 Matches 140; Conservative 94; Mismatches 219; Indels 43; Gaps 13;

OY 4 PALLASALCTFYPLILFLAALKLMDLYCVSSSDRSCALPLPGMTGPFPGFETLQWVLO 63
 DB 20 PALITITIVVVVVVLE-----KMWLMHKEQR-----LRLPGSGMLPYIGETLRYTE 68
 OY 64 R-RKFLQMKRRKKGFTYKTHLFGFPTVRVWGADNVRRIILGSHRLYSVHMPASVFTILGA 122
 DB 69 NPNSEFPTROKNGKIDFKTHILGCPVCMISSPPAARVAVSKAHLKRPYPRPKEMIGP 128
 OY 123 GCLSNLHDSHKOKKRYVMOAFSREALQCYVLIABEVSSCLEQWLSGGERGLLVPEVK 182
 DB 129 EALFFHGGPYHSLKRLVQSSFMPSALRPVSHIELVLOTLSMTSOKSINTLEY--MK 186
 OY 183 RLMRIRMLILGCEPFPAGGEDQQLVEA---PEEMTRNLFSLPIDVPFGILRGVK 238
 DB 187 RYADVAVIM-----SAGDKEEPTTIDYIKLRYLERGNSMPLDLPGLFHKSMK 238
 OY 239 ARNLHARIEENIRAKI--RLIQATEPDGCKDALQLLIEHSWENGERLDMQALKQSTEL 297
 DB 239 AR-----IELSELKRYIEKRENEREG---LLGVLLGAKDQKRNGLSDSQIADNITIGV 291
 OY 298 LFGHETTASANSILTYLGLYPHYLOKVA-REETKSGKLLCKSNQDNKLMETLEQKLT 356
 DB 292 IFATDTTASVLTMLKTYLDHPNMLQEVSRQPSIRQIKIKENR--RISWEDTRKMLPT 349
 OY 357 GCVCYKELRLNMPVPGGERVALKTFEELNGYQIPKGMNVIYICDTHVADITFNKEENP 416
 DB 350 TRVIOETIRASVSTFTRFNAVQVEYDGLIRPKGMVLYPLFRIHSSSEFPPEKFPD 409
 OY 417 DRF-IVPHEDASRFSFLPECGGLRSCVKEFAKILIKITVELARHCDWOLNCPPTMK 475
 DB 410 SREFVAPKRP-----YTYMPFGNGVHSCPGSELAKLEMLILHLHTTFRMEVIGDEBQ 464
 OY 476 TSPVYVPYDNLPAFT 491
 DB 465 YGPFVPRKGLPIRYT 480

RESULT 12
 ID 065624 PRELIMINARY; PRT; 457 AA.
 AC 065624;
 DT 01-AUG-1998 (TREMBLrel. 07, Created)
 DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE CYTOCHROME P450.
 GN T18B16.200 OR AT4G19230.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Magnoliophyta; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 OC Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Bevan M., Benes V., Rechmann S., Borkova D., Ansoerge W., Bancroft I.,
 RA Mayer K.F.X.;
 RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 131-457 FROM N.A.
 RA van der Schueren J., Chuang Y.J., Aert R., Defoor E., Robben J.,
 RA Volckaert G., Mewes H.W., Lemcke K., Mayer K.F.X.;
 RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.

RA EU Arabidopsis sequencing project;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 DR EMBL; AL021687; CAI6713.1; -;
 DR EMBL; AL161550; CAB78925.1; -;
 DR HSSP; P33006; ICPT.
 DR InterPro; IPR001128; -;
 DR Pfam; PF00067; p450.1.
 DR PRINTS; PR00385; P450.
 DR PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_1.
 KW Heme; Monooxygenase; Oxidoreductase.
 SQ SEQUENCE 457 AA; 52436 MW; 6105FB7C181E4F07 CRC64;

Query Match 18.9%; Score 494.5; DB 10; Length 457;
 Best Local Similarity 29.5%; Pred. No. 2,6e-32;
 Matches 137; Conservative 92; Mismatches 191; Indels 45; Gaps 13;

OY 18 LILFLAALKIMDLICVSSRDR--SCALPLPGMTGPFPGFETLQWVLO--RRKFLQMKRRK 74
 DB 8 LTLFASGLFLYFLKRLISRRGSSKRLPLPGMGMPVYGETFOLYSDPNVFPQSKOR 67
 OY 75 YGFYTYTHLFGFPTVRVWGADNVRRIILGSHRLYSVHMPASVFTILGAGCLSNLHDSHK 134
 DB 68 YGSVFETHVLGCPVCMISSPEAKFVLVTKSHLKFPTFASKERMLGKQALFFHGGDYHA 127
 OY 135 QKKVYVMOAFSREALQCYVLIABEVSSCLEQWLSGGERGLLV--YPEVKRLMFRIAMRI 192
 DB 128 KLKRLVLRAMPSESTIRNWPDIESTIADSLRSV-----EGTMINIYQEKYTFPVALIS 182
 OY 193 ILGCEPFPAGGEDE---QQLVEAFEMTRNLFSLPIDVPFGILRGVKANRLHARIE 248
 DB 183 IF-----GKDEVLYRDLKRCYIILEKGYNSMPLDLPGLFHKSMKARREL--SQL 232
 OY 249 ENIRAKIRLIQATEPDGCKDALQLLIEHSWENGERLDMQALKQSTELFGHETTASA 308
 DB 233 ARILSERRQNGSSHND-----LLGSEMGDKLEELTDRO--INDNITIGVIFARDTTAV 283
 OY 309 ATSLITYLGLYPHYLOKVA-REETKSGKLLCKSNQDNKLMETLEQKLTGYICVYKELRLNP 368
 DB 284 MSWIKTYLAENPNVLEAVTE--QMAIRKDKEGESLTMGDTKMKPLTSRVIOETIRVAS 341
 OY 369 PVPQGFVALKTFELNGYQIPKGMNVIYICDTHVADITFNKEENPDRF-IVPHPEDA 427
 DB 342 ILSPTFREAVEDEVEGYLLPKGMKVLPLFRNIHSDIFSNGRKDPSPHFVAPKPN-- 399
 OY 428 SRSFPTPGGLNSCVCKEFAKILIKI---FTVELARHCDWOL 468
 DB 400 ---TFMPFGNGTHSCPGNELAKLEMSIMIHHLTTKRCVCVYTL 441

RESULT 13
 ID 064989 PRELIMINARY; PRT; 513 AA.
 AC 064989;
 DT 01-AUG-1998 (TREMBLrel. 07, Created)
 DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE STEROID 22-ALPHA-HYDROXYLASE.
 GN DMF4.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 OC Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=WS-2;
 RA Choe S., Dilkes B.P., Fujioka S., Takatsuto S., Sakurai A.,
 RA Feldmann K.A.;
 RL Plant Cell 10:0-0(1998).
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 DR EMBL; AF044216; AAC05093.1; -;

DR InterPro: IPR001128; -
 DR Pfam: PF00067; P450; 1.
 DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
 KW Heme: Monooxygenase; Oxidoreductase.
 SQ SEQUENCE 513 AA; 58867 MW; B1639BD9A4DA6F3 CRC64;

Query Match 17.6%; Score 461; DB 10; Length 513;
 Best Local Similarity 28.4%; Pred. No. 1.7e-29;
 Matches 148; Conservative 78; Mismatches 231; Indels 64; Gaps 14;

OY 3 LPALLASLCTFVLPPLFLAALKLMDLYCVSSDRSCALPLPGTGMGFPFGETL----- 58
 DB 10 LPLLLPSSL-----LSLLFLFLK-----RRNRKTRFNLPGKSGWPLGFTIGLK 57
 OY 59 -QWVLRKFLQMKRRKRYGFTKTHLFGPRTVRWGADNVRILLGHRVSVMPASVR 117
 DB 58 PYATTITIGDMQOHVSXGKTYRNSLFGPTVSADAGLNRFILQNGRLFECSYPSIG 117
 OY 118 TILGAGCLSNLHDSHKORKKVIQAFSREALQCYLV-IAEYSSCLEQWLSGGERGL 176
 DB 118 GILGKMSMLVYGDHMDMSISNLFSLHARLRTILKDYERHTLVLDMSQO--NSIFS 175
 OY 177 VYEVKRLMRIRARILLGCEPAGGDEQOLVEAFEEETRLNLSLPIDVPSGLYRG 236
 DB 176 AODEAKFTFNLAHAKHMSMDPGE---EETDLKKEYVTFMKGVASAPLNLPCTAYHKA 231
 OY 237 VKARNLIHARIEENIRAKIRLQATEPDGCKDALQLLIHSWE-----RGRLLMQ- 288
 DB 232 LQSRATILKTEIRKME--RKLDIKEED---QEEREKTEDEAMSKSDHVRKQRTDDL 286
 OY 289 ---ALKOS--STE-----LFGHETTASAATSLITYLGLYPHVLOKVEREIKSKGL 335
 DB 287 LGWVLKHSNSTEQIIDLILSLFAGHETSSVAIALAIFLQACPKAVEELREHLEIAR 346
 OY 336 LCKSNODNKLDMETLEQKTYGIVKETLRLNPVPGFVALKTELVNGYQIPKGNV1 395
 DB 347 AKKELGESSELNMDYKMDTQCYINETLRLGNVVRFLHKKALKDVRYKKYDIPSGKVL 406
 OY 396 YSICDTHDVADIFTNKEEFPDRFIVPHEDASRFS-----FIPFGGGRSCVYKGE 446
 DB 407 PYISAVHLNDSRYDQNLFPWRMOOONNGASSGSGSFSTGNNNMPFGGGRRLCAGSE 466
 OY 447 FAKILKITEVELARHCQDQNLGPRPTMKTSPTVYVPDNL 487
 DB 467 LAKLEMAVFILHLVLKFNWELAE---DDQPPAFPVDP 502

RESULT 14

O9SCQ9
 O9SCQ9 PRELIMINARY; PRT; 513 AA.

AC 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, last annotation update)
 DE STERIOD 22-ALPHA-HYDROXYLASE (DMP4).
 GN T3A5.40.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 OC Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Bloecher H., Mewes H.W., Lemcke K., Mayer K.F.X., Quetler F.,
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 RP [12]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 DR EMBL; AL132979; CAB62435.1; -.

DR InterPro: IPR001128; -
 DR Pfam: PF00067; P450; 1.
 DR PROSITE: PS00385; P450.
 DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
 KW Heme: Monooxygenase; Oxidoreductase.
 SQ SEQUENCE 513 AA; 58867 MW; B1639BD9A5D7C93 CRC64;

Query Match 17.6%; Score 461; DB 10; Length 513;
 Best Local Similarity 28.4%; Pred. No. 1.7e-29;
 Matches 148; Conservative 78; Mismatches 231; Indels 64; Gaps 14;

OY 3 LPALLASLCTFVLPPLFLAALKLMDLYCVSSDRSCALPLPGTGMGFPFGETL----- 58
 DB 10 LPLLLPSSL-----LSLLFLFLK-----RRNRKTRFNLPGKSGWPLGFTIGLK 57
 OY 59 -QWVLRKFLQMKRRKRYGFTKTHLFGPRTVRWGADNVRILLGHRVSVMPASVR 117
 DB 58 PYATTITIGDMQOHVSXGKTYRNSLFGPTVSADAGLNRFILQNGRLFECSYPSIG 117
 OY 118 TILGAGCLSNLHDSHKORKKVIQAFSREALQCYLV-IAEYSSCLEQWLSGGERGL 176
 DB 118 GILGKMSMLVYGDHMDMSISNLFSLHARLRTILKDYERHTLVLDMSQO--NSIFS 175
 OY 177 VYEVKRLMRIRARILLGCEPAGGDEQOLVEAFEEETRLNLSLPIDVPSGLYRG 236
 DB 176 AODEAKFTFNLAHAKHMSMDPGE---EETDLKKEYVTFMKGVASAPLNLPCTAYHKA 231
 OY 237 VKARNLIHARIEENIRAKIRLQATEPDGCKDALQLLIHSWE-----RGRLLMQ- 288
 DB 232 LQSRATILKTEIRKME--RKLDIKEED---QEEREKTEDEAMSKSDHVRKQRTDDL 286
 OY 289 ---ALKOS--STE-----LFGHETTASAATSLITYLGLYPHVLOKVEREIKSKGL 335
 DB 287 LGWVLKHSNSTEQIIDLILSLFAGHETSSVAIALAIFLQACPKAVEELREHLEIAR 346
 OY 336 LCKSNODNKLDMETLEQKTYGIVKETLRLNPVPGFVALKTELVNGYQIPKGNV1 395
 DB 347 AKKELGESSELNMDYKMDTQCYINETLRLGNVVRFLHKKALKDVRYKKYDIPSGKVL 406
 OY 396 YSICDTHDVADIFTNKEEFPDRFIVPHEDASRFS-----FIPFGGGRSCVYKGE 446
 DB 407 PYISAVHLNDSRYDQNLFPWRMOOONNGASSGSGSFSTGNNNMPFGGGRRLCAGSE 466
 OY 447 FAKILKITEVELARHCQDQNLGPRPTMKTSPTVYVPDNL 487
 DB 467 LAKLEMAVFILHLVLKFNWELAE---DDQPPAFPVDP 502

RESULT 15

O9FNAS
 O9FNAS PRELIMINARY; PRT; 465 AA.

AC 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, last annotation update)
 DE CYTOCHROME P450.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 OC Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Sato S., Kaneko T., Kotani H., Nakamura Y., Asamizu E., Miyajima N.,
 RL MEDLINE=98290546; PubMed=9628582;
 RA Tabata S.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. IV.
 RT Sequence features of the regions of 1,456,315 bp covered by nineteen
 RT physically assigned P1 and TAC clones."
 RL DNA Res. 5:41-54(1998)
 DR EMBL; AB009048; BAB08653.1; -.

SEQ SEQUENCE 465 AA; 53767 MW; 0C00459C9C866D1F CRC64;

Query Match 17.48; Score 455.5; DR 10; Length 465;

Best Local Similarity 28.28; Pred. No. 4.1e-29;

Matches 137; Conservative 91; Mismatches 211; Indels 47; Gaps 13;

```

QY 1 MGLPALASALCTFVPLLLFLAALKIMDLYCVSSRDRSCALPLPQTGPFPPFGETLQM 60
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
Db 8 MGLLLIIVS-LC-----SALLRW-----NOMRYTKNGLPCTGMWPPIGETTEF 50
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
QY 61 VLORRRKFLQKRRKRYGYFYKTHLEGRPTVYVNGADNVRRILLGEHRLVSVWPAVPTIL 120
   :||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
Db 51 LKGPDMFMNRQRLRYGSFPRSHLLGCTPLISMSEVNRITLKNSKGLVPQYDQSMIDL 110
   :||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
QY 121 GAGCLSNLHDSHKOKKKYIMQAFSREALOCYLVIAEF-VSSCLEQWMLSCGERGLL-VY 178
   :||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
Db 111 GTCNMAVHSGSHRLMKGSLSLISSTMMDHILPKVDHFMRSYLDQW--NELEVIDIQ 167
   :||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
QY 179 PEYKRLMFRFAMRILLGCEHPGAGGEDQOLVEAFEEEMTRNLFSLPIDVPFSGLYRGVK 238
   :||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
Db 168 DKTKHMAFLSSLQIAGNLKKP-----VEEKTAFPKLVVGTLSVPIDLPGTNYRCGIQ 222
   :||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
QY 239 ARNLIHARIEENTIRAKIRRLQATEPDG--CKDALQLLIEHSWEGERLDMQALKOSTE 296
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
Db 223 ARN-----NIDRLRELMQERROSGEFTMDLGYLMKKEGNRYPLTD-EIIRDQVVT 273
   :||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
QY 297 LTFGHEHTASAAFTSLITYGLYPHYLQKVREIKSKGLLCKSNQDNKLDMETLEQLKYI 356
   :||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
Db 274 ILYGYETVSTSMALKYLDHPKALQELRAE--HLAFREKRKRODEPLGLEEDVKSMTFT 331
   :||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
QY 357 GCYIKETLRNPPVPGGFRALKTFFELNGYQIPKGMNVIYSICDTHVADIFTNKEENP 416
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
Db 332 RAVIYETSRLATVNGVLRKTRTDLEINGYLIPKGMRIYVYTRREINDANLYEDPLIFNP 391
   :||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
QY 417 DRFTVPHPEDASRFSPFPGGLRSCVYGEFAKILKIFTEIARHCDWQJLNGPPTMKT 476
   :||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
Db 392 WRMMKKSLE--SQNSCFVFGGTRLCPRKELGIVEISSFLHYFVTRYRWEIIGGDELM-- 447
   :||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
QY 477 SPTVYP 482
   :||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
Db 448 ---VFP 450

```

Search completed: November 6, 2001, 13:38:28
Job time: 256 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2001, 13:39:40 ; Search time 62.8 Seconds
(without alignments)
271.098 Million cell updates/sec

Title: US-09-668-482-32

Perfect score: 2612
Sequence: 1 MGLPALLASALCTFVLPDLL.....PTYPVNDLPARTYFGDI 497

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2612	100.0	497	1 CP26_MOUSE	055127 mus musculu
2	2441	93.5	497	1 CP26_HUMAN	043174 homo sapien
3	1734	66.4	492	1 CP26_BRARE	P79739 brachydanto
4	450	17.2	472	1 CP91_ARATH	042569 arabidopsis
5	442	16.9	464	1 CP85_LYCES	043147 lycopersico
6	442	16.9	490	1 CP83_ARATH	023051 arabidopsis
7	410	15.7	492	1 C136_MYCTU	P95099 mycobacteri
8	359	13.7	519	1 CP81_MAIZE	043246 zea mays (m
9	339	13.0	503	1 CP51_RAT	064654 rattus norv
10	337	12.9	520	1 CP84_ARATH	042600 arabidopsis
11	332	12.7	503	1 CP51_PIG	046420 sus scrofa
12	327.5	12.5	505	1 CP48_DROME	09vs79 drosophila
13	326.5	12.5	493	1 C132_DROME	09v9p4 drosophila
14	326	12.5	501	1 CP42_DROME	027589 drosophila
15	326	12.5	503	1 CP51_HUMAN	016850 homo sapien
16	325	12.4	511	1 CP4C_BIADI	P29981 blaberus di
17	323	12.4	472	1 C13B_MYCTU	053755 mycobacteri
18	322.5	12.3	476	1 C137_MYCTU	069653 mycobacteri
19	322.5	12.3	511	1 CP45_RABIT	P14579 oryctolagus
20	317	12.1	441	1 C138_MYCTU	P96813 mycobacteri
21	314.5	12.0	512	1 C4D1_DROST	P33269 drosophila
22	312	11.9	518	1 CP3R_ONCMY	042563 drosophila
23	309	11.8	494	1 C134_DROME	09v9d0 drosophila
24	308.5	11.8	459	1 CPXN_AKNSP	P29960 anabena sp
25	307.5	11.8	511	1 CP47_RABIT	P14581 oryctolagus
26	307	11.8	483	1 C311_DROME	09v9t7 drosophila
27	307	11.8	511	1 C4AE_DROME	046054 drosophila
28	307	11.8	511	1 C772_SOIME	P37124 solanum mel
29	305.5	11.7	512	1 C4D1_DROST	016805 drosophila
30	305.5	11.7	519	1 CP4T_HUMAN	002928 homo sapien
31	305	11.7	500	1 CP46_MOUSE	09v9v8 mus musculu
32	304.5	11.7	510	1 CP46_RABIT	P14580 oryctolagus
33	301.5	11.5	487	1 C135_DROME	09v9p5 drosophila

34	301	11.5	520	1 CP33_HUMAN	008477 homo sapien
35	300.5	11.5	537	1 CP66_RAT	P51871 rattus norv
36	300	11.5	576	1 C972_SOYBN	048921 glycine max
37	299.5	11.5	509	1 CP41_RAT	P08516 rattus norv
38	298	11.4	500	1 CP46_HUMAN	09v6a2 homo sapien
39	296.5	11.4	507	1 CP3S_BOVIN	P79102 bos taurus
40	293.5	11.2	512	1 CP51_CUNEL	09v9c3 cunnilingame
41	293	11.2	506	1 CP44_RABIT	P10611 oryctolagus
42	293	11.2	507	1 C4DE_DROME	046051 drosophila
43	291.5	11.2	508	1 CP48_RAT	P24464 rattus norv
44	289.5	11.1	492	1 C133_DROME	09v9b3 drosophila
45	289.5	11.1	506	1 CP4B_RABIT	P15128 oryctolagus

ALIGNMENTS

RESULT 1
ID CP26_MOUSE STANDARD: PRT; 497 AA.
AC 055127: CP26_MOUSE
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE CYTOCHROME P450 26 (EC 1.14.-.-) (RETINOIC ACID-METABOLIZING
DE CYTOCHROME) (P450RA1) (RETINOIC ACID 4-HYDROXYLASE).
GN CYP26A1 OR CYP26 OR P450RA.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C3H;
RX MEDLINE=97392446; PubMed=9250660;
RA Fujii H., Sato T., Kaneko S., Gotoh O., Fujii-Kuriyama Y., Osawa K.,
RA Kato S., Hamada H.;
RT "Metabolic inactivation of retinoic acid by a novel P450
RL EMBO J. 16:4163-4173(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=98113212; PubMed=9442090;
RA Abu-Abed S.S., Beckett B.R., Chiba H., Chithalen J.V., Jones G.,
RA Metzger D., Chambon P., Petkovich M.;
RT "Mouse P450RA1 (CYP26) expression and retinoic acid-inducible retinoic
RT acid metabolism in F9 cells are regulated by retinoic acid receptor
RT gamma and retinoid X receptor alpha.";
RL J. Biol. Chem. 273:2409-2415(1998).
CC -1- FUNCTION: PLAYS A KEY ROLE IN RETINOIC ACID METABOLISM. ACTS ON
CC RETINOIDS, INCLUDING ALL-TRANS-RETINOIC ACID (RA) AND ITS
CC STEREOISOMER 9-CIS-RA. CAPABLE OF BOTH 4-HYDROXYLATION AND 18-
CC HYDROXYLATION. RESPONSIBLE FOR GENERATION OF SEVERAL HYDROXYLATED
CC FORMS OF RA, INCLUDING 4-OH-RA, 4-OXO-RA, AND 18-OH-RA.
CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM.
CC -1- INDUCTION: BY RETINOIC ACIDS (RA).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-slb.ch/announce/>
CC or send an email to license@isb-slb.ch).
CC
CC EMBL: Y12657; CAA73206.1; -
CC MGD: MGI:1096359; Cyp26.
CC InterPro: IPR001128; -
CC Pfam: PF00067; P450; 1.
CC PRINTS: PR00385; P450.
CC PROSITE: PS00086; CYTOCHROME_P450; 1.
CC OXIDOREDUCTASE; Monooxygenase; Membrane; Heme; Microsome;

KW Endoplasmic reticulum.
FT BINDING 442 442 HEME (POTENTIAL).
SQ SEQUENCE 497 AA; 56177 MW; 33807D/C29134471 CRC64;

Query Match 100.0%; Score 2612; DB 1; Length 497;
Best Local Similarity 100.0%; Pred. No. 5,7e-175;
Matches 497; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
OY 1 MGIPALLASALCTFVPLPILFLAALKLMDLYCVSSNRSCALPLPPTMGPFPGETLQ 60
DB 1 MGIPALLASALCTFVPLPILFLAALKLMDLYCVSSNRSCALPLPPTMGPFPGETLQ 60
OY 61 VIORRRFLQMKRRKRGFYIKTHLFGRPYRVMGADVNRILLGEHRLVSHWPAVSRTIL 120
DB 61 VIORRRFLQMKRRKRGFYIKTHLFGRPYRVMGADVNRILLGEHRLVSHWPAVSRTIL 120
OY 121 GAGCISNLHDSHKKRKRYIMQAFSREALQCYLVIAEEVSSCLEQWLSGGERGLLYPE 180
DB 121 GAGCISNLHDSHKKRKRYIMQAFSREALQCYLVIAEEVSSCLEQWLSGGERGLLYPE 180
OY 181 VKRLMFRIMRILLGCEPPAGGGEEDQOLVEAFEMTRNLSLPIDVPFSGLYRGYKAR 240
DB 181 VKRLMFRIMRILLGCEPPAGGGEEDQOLVEAFEMTRNLSLPIDVPFSGLYRGYKAR 240
OY 241 NLIHARIEENIRAKIRRLQATEPDDGCKDALQLLIEHSMERGEKRLDMQALKOSTELLFG 300
DB 241 NLIHARIEENIRAKIRRLQATEPDDGCKDALQLLIEHSMERGEKRLDMQALKOSTELLFG 300
OY 301 GHETTASATSLITTYLGLYVHLYQKRYREIKSKGLCKSNQDKLDMETLEQIKYICVYI 360
DB 301 GHETTASATSLITTYLGLYVHLYQKRYREIKSKGLCKSNQDKLDMETLEQIKYICVYI 360
OY 361 KETLRINPVPVGGFRVALKTFELNGYQIIPKGNVYISICDTHDVADITNKEEFNDRFI 420
DB 361 KETLRINPVPVGGFRVALKTFELNGYQIIPKGNVYISICDTHDVADITNKEEFNDRFI 420
OY 421 VHPEDASFSFIPGGGGRVSCVGRKFAKILKIFVELARHCDQLLNGPPIMKTSPIVY 480
DB 421 VHPEDASFSFIPGGGGRVSCVGRKFAKILKIFVELARHCDQLLNGPPIMKTSPIVY 480
OY 481 YPVDNLPAFTYFGDI 497
DB 481 YPVDNLPAFTYFGDI 497
```

RESULT 2
CP26_HUMAN STANDARD; PRT; 497 AA.
ID CP26_HUMAN 043174;
AC 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE CYTOCHROME P450 26 (EC 1.14.-.-) (RETINOIC ACID-METABOLIZING
CYTOCHROME) (P450RA1) (RETINOIC ACID 4-HYDROXYLASE).
GN CYP26A1 OR CYP26.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9373542; PubMed=9228017;
RA White J.A., Beckett-Jones B., Guo Y.-D., Dilworth F.J., Bonasoro J.,
RA Jones G., Petkovich M.;
RT "cDNA cloning of human retinoic acid-metabolizing enzyme (hp450RA1)
RT identifies a novel family of cytochrome P450.";
RL J. Biol. Chem. 272:18538-18541(1997);
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=98380037; PubMed=9716180;
RA Soneveld E., van den Brink C.E., van der Lee de B.M., Schultes R.K.,
RA Petkovich M., van der Burg B., van der Saag P.T.;

RT "Human retinoic acid (RA) 4-hydroxylase (CYP26) is highly specific for
RT all-trans-RA and can be induced through RA receptors in human breast
RT and colon carcinoma cells."
RL Cell Growth Differ. 9:629-637(1998).
RN [3]
RP TISSUE SPECIFICITY.
RX MEDLINE=99045433; PubMed=9826557;
RA Trofimova-Griffin M.E., Juchau M.R.;
RT "Expression of cytochrome P450RA1 (CYP26) in human fetal hepatic and
RT cephalic tissues."
RL Biochem. Biophys. Res. Commun. 252:487-491(1998).
CC -1- FUNCTION: PLAYS A KEY ROLE IN RETINOIC ACID METABOLISM. ACTS ON
CC RETINOIDS, INCLUDING ALL-TRANS-RETINOIC ACID (RA) AND ITS
CC STEREOISOMER 9-CIS-RA. CAPABLE OF BOTH 4-HYDROXYLATION AND 18-
CC HYDROXYLATION. RESPONSIBLE FOR GENERATION OF SEVERAL HYDROXYLATED
CC FORMS OF RA, INCLUDING 4-OH-RA, 4-OXO-RA, AND 18-OH-RA.
CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM.
CC -1- TISSUE SPECIFICITY: HIGHEST LEVELS IN ADULT LIVER, HEART,
CC PITUITARY GLAND, ADRENAL GLAND, PLACENTA AND REGIONS OF THE BRAIN.
CC -1- INDUCTION: BY RETINOIC ACIDS (RA).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF005418; AAB88881.1; -.
DR MIM; 602239; -.
DR InterPro; IP001128; -.
DR Pfam; PF00067; P450; 1.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
KW Oxidoreductase; Monooxygenase; Membrane; Heme; Microsome;
KW Endoplasmic reticulum.
FT BINDING 442 442 HEME (POTENTIAL).
SQ SEQUENCE 497 AA; 56162 MW; EAB6B44B2AB2EAB3 CRC64;

Query Match 93.5%; Score 2441; DB 1; Length 497;
Best Local Similarity 93.4%; Pred. No. 4,7e-163;
Matches 464; Conservative 16; Mismatches 17; Indels 0; Gaps 0;

```
OY 1 MGIPALLASALCTFVPLPILFLAALKLMDLYCVSSNRSCALPLPPTMGPFPGETLQ 60
DB 1 MGIPALLASALCTFVPLPILFLAALKLMDLYCVSSNRSCALPLPPTMGPFPGETLQ 60
OY 61 VIORRRFLQMKRRKRGFYIKTHLFGRPYRVMGADVNRILLGEHRLVSHWPAVSRTIL 120
DB 61 VIORRRFLQMKRRKRGFYIKTHLFGRPYRVMGADVNRILLGEHRLVSHWPAVSRTIL 120
OY 61 VIORRRFLQMKRRKRGFYIKTHLFGRPYRVMGADVNRILLGEHRLVSHWPAVSRTIL 120
DB 61 VIORRRFLQMKRRKRGFYIKTHLFGRPYRVMGADVNRILLGEHRLVSHWPAVSRTIL 120
OY 121 GAGCISNLHDSHKKRKRYIMQAFSREALQCYLVIAEEVSSCLEQWLSGGERGLLYPE 180
DB 121 GAGCISNLHDSHKKRKRYIMQAFSREALQCYLVIAEEVSSCLEQWLSGGERGLLYPE 180
OY 121 GAGCISNLHDSHKKRKRYIMQAFSREALQCYLVIAEEVSSCLEQWLSGGERGLLYPE 180
DB 121 GAGCISNLHDSHKKRKRYIMQAFSREALQCYLVIAEEVSSCLEQWLSGGERGLLYPE 180
OY 181 VKRLMFRIMRILLGCEPPAGGGEEDQOLVEAFEMTRNLSLPIDVPFSGLYRGYKAR 240
DB 181 VKRLMFRIMRILLGCEPPAGGGEEDQOLVEAFEMTRNLSLPIDVPFSGLYRGYKAR 240
OY 181 VKRLMFRIMRILLGCEPPAGGGEEDQOLVEAFEMTRNLSLPIDVPFSGLYRGYKAR 240
DB 181 VKRLMFRIMRILLGCEPPAGGGEEDQOLVEAFEMTRNLSLPIDVPFSGLYRGYKAR 240
OY 241 NLIHARIEENIRAKIRRLQATEPDDGCKDALQLLIEHSMERGEKRLDMQALKOSTELLFG 300
DB 241 NLIHARIEENIRAKIRRLQATEPDDGCKDALQLLIEHSMERGEKRLDMQALKOSTELLFG 300
OY 241 NLIHARIEENIRAKIRRLQATEPDDGCKDALQLLIEHSMERGEKRLDMQALKOSTELLFG 300
DB 241 NLIHARIEENIRAKIRRLQATEPDDGCKDALQLLIEHSMERGEKRLDMQALKOSTELLFG 300
OY 301 GHETTASATSLITTYLGLYVHLYQKRYREIKSKGLCKSNQDKLDMETLEQIKYICVYI 360
DB 301 GHETTASATSLITTYLGLYVHLYQKRYREIKSKGLCKSNQDKLDMETLEQIKYICVYI 360
OY 301 GHETTASATSLITTYLGLYVHLYQKRYREIKSKGLCKSNQDKLDMETLEQIKYICVYI 360
DB 301 GHETTASATSLITTYLGLYVHLYQKRYREIKSKGLCKSNQDKLDMETLEQIKYICVYI 360
OY 361 KETLRINPVPVGGFRVALKTFELNGYQIIPKGNVYISICDTHDVADITNKEEFNDRFI 420
DB 361 KETLRINPVPVGGFRVALKTFELNGYQIIPKGNVYISICDTHDVADITNKEEFNDRFI 420
```

```

QY 421 VPHEDASRFSFIPFGGGLRSCVGEKFAKILIKFTVELARHCMDQLNGPPTMKTSPV 480
DB 421 LPHEDASRFSFIPFGGGLRSCVGEKFAKILIKFTVELARHCMDQLNGPPTMKTSPV 480
QY 481 YPVNDLPARFTYFQGD 497
DB 481 YPVNDLPARFTYFQGEI 497

RESULT 3
CP26_BRAVE STANDARD: PRT: 492 AA.
AC P79739;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE CYTOCHROME P450 26 (EC 1.14.-.-) (RETINOIC ACID-METABOLIZING
DE CYTOCHROME) (P450RA1) (RETINOIC ACID 4-HYDROXYLASE).
CN CYP26.
OS Brachydanio rerio (Zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC Cypriniformes; Cyprinidae; Rasbora; Danio.
OX NCBI_Taxid=7955;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97094702; PubMed=8939936;
RA White J.A., Guo Y.-D., Baetz K., Beckett-Jones B., Bonasoro J.,
RA Hsu K.E., Dilworth F.J., Jones G., Petkovich M.;
RT "Identification of the retinoic acid-inducible all-trans-retinoic
RT acid 4-hydroxylase."
RU J. Biol. Chem. 271:29922-29927(1996).
CC -1- FUNCTION: PLAYS A KEY ROLE IN RETINOIC ACID METABOLISM. ACTS ON
CC RETINOIDS, INCLUDING ALL-TRANS-RETINOIC ACID (RA) AND ITS
CC STEREOISOMER 9-CIS-RA. CAPABLE OF 4-HYDROXYLATION, HYDROXYLATION.
CC RESPONSIBLE FOR GENERATION OF SEVERAL HYDROXYLATED FORMS OF RA,
CC INCLUDING 4-OH-RA AND 4-OXO-RA.
CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM.
CC -1- INDUCTION: BY RETINOIC ACIDS (RA).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U68234; AAC60045.1; -.
DR ZFIN: ZDB-GENE-990415-44; cyp26.
DR InterPro: IPR001128; -.
DR Pfam: PF00067; P450.1.
DR PRINTS: PR00385; P450.
DR PROSITE: PS00086; CYTOCHROME_P450.1.
KW Oxidoreductase; Monooxygenase; Membrane; Heme; Microsome;
KW Endoplasmic reticulum.
FT BINDING 438 438 HEME (BY SIMILARITY).
SQ SEQUENCE 492 AA; 56281 MW; FD471435B2F30509 CMC64;

Query Match 66.4%; Score 1734; DB 1; Length 492;
Best Local Similarity 67.7%; Pred. No. 8.6e-114;
Matches 336; Conservative 61; Mismatches 89; Indels 10; Gaps 4;

```

```

QY 121 GAGCLSNLHDSHKKRKVIMQAFREALQCYVLYIAEVSQCLEQWL---SCGREGILV 177
DB 121 GSDLTSLNVHGVQKKKKRAIMAFESRDALERTIPVQDEVSAIDQWLOKQSC---VLV 176
QY 178 YPEVKRLMFRIAMRILLCEPAGGEDQOLVAFEMTNLFLPIDVFFSGLRGV 237
DB 177 YPEKKLMFRIAMRILLQFEPQI--KTDEQLVAFAFEMIKNLFSLPIDVFFSGLRGL 234
QY 238 KARNLIHARIEENIRAKIRLQATEPDGCKDALQLLIHSEMEREDMDQAKOSSTEL 297
DB 235 KARNFIHSHKIEENIRKIKIOD--DNNENQKYKDALQLLIENSRSDEPSLOAKMEATEL 293
QY 298 LFGGHETASATSLITLGLVPHYLQVREIRKSGLLCKSNQNMKMDMETLDELQKYG 357
DB 294 LFGGHETASATSLVLMELGLTEVQVREVRQREVMGMYTPGKGLSMETLDELQKYG 353
QY 358 CVIKETLRNPPVPGGFRAVAKTFELNGVQIPKGMNVYISICDTHDVAIDFTNKEEFND 417
DB 354 CVIKETLRNPPVPGGFRAVAKTFELNGVQIPKGMNVYISICDTHDVAIDFTNKEEFQPE 413
QY 418 RYVPHEDASRFSFIPFGGGLRSCVGEKFAKILIKFTVELARHCMDQLNGPPTMKT 477
DB 414 RYMSKGLDEGSFENYIPFGGSRMCVGEKFAKILIKFTVELARHCMDQLNGPPTMKT 473
QY 478 PTVYVNDLPARFTYF 493
DB 474 PTVYVNDLPARFTYF 489

```

```

RESULT 4
C90L_ARATH STANDARD: PRT: 472 AA.
ID C90L_ARATH
AC Q42563;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE 15-DEC-1998 (Rel. 37, Last annotation update)
DE CYTOCHROME P450 90A1 (EC 1.14.-.-).
CN CYP90A1 OR CYP90 OR CPD.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_Taxid=3702;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=CV. COLUMBIA;
RX MEDLINE=96200769; PubMed=8612270;
RA Szekeres M., Nemeth K., Koncz-Kalman Z., Mathur J., Kauschmann A.,
RA Altman T., Redei G.P., Nagy F., Schell J., Koncz G.;
RT "Brassinosteroids rescue the deficiency of CYP90, a cytochrome P450,
RT controlling cell elongation and de-etiolation in Arabidopsis."
RL Cell 85:171-182(1996).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X87367; CAA60793.1; -.
DR EMBL: X87368; CAA60794.1; -.
DR InterPro: IPR001128; -.
DR Pfam: PF00067; P450.2.
DR PROSITE: PS00086; CYTOCHROME_P450.1.
KW Oxidoreductase; Monooxygenase; Heme.
FT BINDING 418 418 HEME (BY SIMILARITY).
SQ SEQUENCE 472 AA; 53785 MW; 41A73F4D64B343F CRC64;

```

```

Query Match      17.2%; Score 450; DB 1; Length 472;
Best Local Similarity 26.6%; Pred. No. 2.5e-24;
Matches 134; Conservatve 90; Mismatches 217; Indels 62; Gaps 15;

QY 6 LLASALCTEVLPLLLFLAALKIMDLVYSSDRSCALPLPGTGMFPFPGTLOWY---- 61
D 9 LLSIAAGFLL-----RRRYRRMGLPGSLGLPLIGETFOIGAYK 52
QY 62 -LQRRKFLQMKRRKRYGFIYKTHLGRPTVRVMGADNVRRIILGHRILVSMIPASVRILL 120
D 53 TENPEPFLDENVARVAGSVFTHLFGEPFLFSADPEPTNRVLQNEKLEECSPASICML 112
QY 121 GAGCISLHDSHKKORRVIMQAFSREALQCYLV-IAEEVSSCLEQWLSGERSLIVYP 179
D 113 GKHSLLMKKSLHRMSLMTSPANSSIIKHMLMDIDRLVRFNLDSMS----RVLLME 168
QY 180 EVKILMERIAMRIILGCEPPAGGEEOQLVEAFEMTRNLFSLPIDVPSGLYR-GVK 238
D 169 EAKRTTELIVKQLMSFDP-----GEMSESIRKEYLIVIEGFSLPLPL-FSTYRKAIQ 222
QY 239 ARNLIHRIEENIRAKIRIQLATEPDG--CKDALQLII--EHSWERGERLDMQALKSS 294
D 223 ARR-----KVALVYVVMKREEBEGERRKMDLAALLADGFSDEIVDF-----L 272
QY 295 TELFGHETTSASATSLITLYGLYPHYLQKREIKSGLLCKSNODNKLDMETLEOLK 354
D 273 VALLVAGETSTIMTLAVKFLTEPLALQLEE--HEKIRAKMSDSYSLIEMSTYKMP 330
QY 355 YIGCVIKETLPLPPVPGGFRVALKTEFLNGYQIPKGMNVITYSCDTHVDADIFTNKEEF 414
D 331 FTQCVNETLLEVANIIGVFRRAMTDEIKYKIPKMKVSSFAVHLADNHFEDARTF 390
QY 415 NPDEF-----IYHPEDASRFSGIFPGGGLRSCGKEPAKILIKIPTVELARHCQMLNG 470
D 391 NPMWQMSNSVTGGSNV-----FTPPGGGPRCLCPGYELARVALSVFLMLVGFSSW----- 441
QY 471 PPTMKISPTVYPVDNLPAFTYF 493
D 442 VPAEDKLVEFPPTRTQKRYPIF 464

RESULT 5
CP85_LYCES STANDARD; PRT; 464 AA.
AC 043147;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE CYTOCHROME P450 85 (EC 1.14.--) (DWARF PROTEIN).
GN CYP85 OR D.
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Asteridae I;
OC Solanales; Solanaceae; Solanum.
OX NCBI_Taxid:4081;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. GCR758;
RX MEDLINE=96266705; Pubmed=8672892;
RA Bishop G.J., Harrison K., Jones J.G.D.;
RT "The tomato Dwarf gene isolated by heterologous transposon tagging
encodes the first member of a new cytochrome P450 family.";
RL Plant Cell 8:959-969(1996).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

DR EMBL: U54770; AAB17070.1; -.
DR InterPro: IPR001128; -.
DR Pfam: PF00067; P450.1.
DR PRINTS: PR00385; P450.
DR PROSITE: PS00086; CYTOCHROME_P450.1.
KW Oxidoreductase; Monooxygenase; Membrane; Heme.
FT BINDING 414 414 HEME (BY SIMILARITY).
SQ SEQUENCE 464 AA; 53706 MW; D2B21AAB7B14E94 CMC64;

Query Match      16.9%; Score 442; DB 1; Length 464;
Best Local Similarity 27.0%; Pred. No. 8.7e-24;
Matches 125; Conservatve 94; Mismatches 210; Indels 34; Gaps 10;

QY 18 LILPLAALKIMDLVYSSDRSCALPLPGTGMFPFPGTLOWVQRRKFLQMKRRKYG 77
D 14 LCIFCTALLRNQVYKNK-----LPPGTMGWPLFGETTEFLIGFSFKNQARSGS 67
QY 78 IYKTHLFRPTVRVMGADNVRRIILGHRILVSMIPASVRITIGACSLNHDSSHKOR 137
D 68 FPKSHILGCPITVSDSLNRIILVNEAKGLVPGYQSMIDILGKCNIAANGSAHKYMR 127
QY 138 KVMQAFSREALQCYLV-IAEEVSSCLEQWLSGERSLIVPEYKRLMERIAMRIILGC 196
D 128 GALLSLISPTMIRDLPLKIDEFMRSHLTNW--DNKVIDIQETKNKAFSLSKQIAGI 184
QY 197 EPPAGGEDEOQLVEAFEMTRNLFSLPIDVPSGLYRKYAR-----NLHRIEENIR 252
D 185 ESTSLA-----QEFMSSEFNVLGTLSPILM.PNNYHGRGQARKIIVNLRTLIEE--R 237
QY 253 AKIRIQLATEPDGCKDMLQILIEHSWERGERLDMQALKQSTELLEFGHETTSASATSL 312
D 238 RASKEIQ-----HMDLGLTMEDEATRRFKLTDEMIDLIT-ILYGYETVSTSSMA 288
QY 313 ITYGLYPHYLQKREIKSGLLCKSNODNKLDMETLEQLKIGCVIKETLRLNPPVPG 372
D 289 VKYLDHPRKVLLELRKE--HMAIRKKKPEDPIDYNDYRSRMRFTRAVILFETSRATIYNG 346
QY 373 GFRVALKTEFLNGYQIPKGMNVITYSCDTHVDADIFTNKEEFNDRFVPPEDASRFSF 432
D 347 VLKRTQDMELNGIILIPGMRILYVYTRRLANDPRLIPPYPYFNMRKMDKSLERON--SF 404
QY 433 IPFGGGLRSCGKEPAKILIKIPTVELARHCQMLNPPPTK 475
D 405 LVFGGTRQCPGKELGVAEISTFLHYETKYRWEIGDKIMK 447

RESULT 6
C883_ARATH STANDARD; PRT; 490 AA.
AC 023051;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE CYTOCHROME P450 88A3 (EC 1.14.--).
GN CYP88A3 OR YUP8H12.23.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_Taxid:3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. COLUMBIA.
RA Theologis A., Osborne B.I., Vysotskaia V.S., Federspiel N.A.,
RA Totlunt M., Yu G., Ojl O., Araujo R., Chung E., Dewar K., Dietrich F.,
RA Ecker J.R., Marzilli A., Oelner P., Davis R.W.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-
CC the European Bioinformatics Institute. There are no restrictions on its

```


Db 478 MIPMDGMP 486

RESULT 8
C881_MAIZE
ID C881_MAIZE STANDARD: PRT: 519 AA.
AC 043246;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE CYTOCHROME P450 88A1 (EC 1.14.-.-) (DMARF3 PROTEIN).
GN CYP88A1 OR D3.
OS Zea mays (maize).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae;
OC Andropogoneae; Zea.
OX NCBI_TaxId=4577;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. B73;
RX MEDLINE=96004534; PubMed=7549486;
RA Winkler R.G., Helentjaris T.;
RT "The maize Dwarf3 gene encodes a cytochrome P450-mediated early step
in Gibberellin biosynthesis."
RL Plant Cell 7:1307-1317(1995).
CC -1- PATHWAY: EARLY STEP IN GIBBERELLIN BIOSYNTHESIS.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN ROOTS, DEVELOPING LEAVES, THE
CC VEGETATIVE MERISTEM, AND SUSPENSION CULTURE CELLS.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U32579; AAC49067.1; -
DR InterPro: IPR001128; -
DR Pfam: PF000067; P450; 1.
DR PROSITE: PS00086; CYTOCHROME_P450; 1.
DR Oxidoreductase; Monooxygenase; Transmembrane; Heme.
KM TRANSMEM
FT BINDING 1 21 POTENTIAL.
FT HEME (BY SIMILARITY).
SQ SEQUENCE 519 AA; 57906 MW; 0F8977A024316D95 CRC64;

Query Match 13.7%; Score 359; DB 1; Length 519;
Best Local Similarity 26.3%; Pred. No. 6.1e-18;
Matches 123; Conservative 89; Mismatches 213; Indels 42; Gaps 15;

QY 44 LPPGWMGPFEGEETLQWLOVRK---FLQMKRRRYG--FTYKTHLFGRPVRYVGAUN 96
DB 72 LPPGWMGPFEGEETLQWLOVRK---FLQMKRRRYG--FTYKTHLFGRPVRYVGAUN 96
QY 97 VRRILGELRLVSVHPASVTRITLGAACLSNLDSSHOKKRVYMOAER-ELAOCTYLV 155
DB 132 CKQVLMDDDAFT-GMPATYALVGRSFVAPYDEHRRIRRLTAAPINGFALTGYLDF 190
QY 156 IAEVSSCLEQWLSGCEGLLVY-PEVKRLMFRIMRILGCEPGPAGGEDE---QOCLV 211
DB 191 IDRTVSSLRAMADHG--GSVEFLRLRMFTKIIVQFL-----GGADQATTRALE 240
QY 212 EAFEMTNNLESPLDVPFSGIYRG-VARNLIHARIEENIRAKIRLQATEPDGCKDA 270
DB 241 RSYTELNGMAMALNIP-GRAYRGALRRARRLVA-VLQGVLDERRARRAKVSGGVDM 298
QY 271 LQLLIHEHMERGERLDMQALQSSSTELLFGGHETASATSLITLYGLVPHYLQVREI 330
DB 299 MDRLLIEADDEGRHLDDEIIDLVMYINAGHESGHITTMATVFLQENPDMFAAKAEQ 358
QY 331 KSKGLLCKSNODNKIMLMTLEBOLKXVIGCVIKRETLRLNPVVGGFVALKTELNGYQIPK 390

Db 359 EAIMRSTPSSO-RGLTIRDFRKMEXLSOVIDETLRNLVNSVSRQATRDVFNGLIPK 417
QY 391 GWNVYISICDTHVDADIFTNKEENPDREIVPHREDASRFSFIPGGGLRSCVKEPAKI 450
DB 418 GWKQVLMYRSVHMDPOYVPPDPTKDFPSRMECHSPRAG---TELFAGGARLCCPNDLAKL 474
QY 451 LKRIET-----VELARHCWQQLNGPPTMTKTSPTVYVYDNLPRFT 491
DB 475 EISVFLHHLGLGYRLAR-----TNPRCVRKYLPHPRPVNCLAKIT 515

RESULT 9
CPS1_RAT
ID CPS1_RAT STANDARD: PRT: 503 AA.
AC 064654; 064549;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE CYTOCHROME P450 51 (EC 1.14.14.1) (CYP11) (P450L1) (STEROL 14-ALPHA
DE DEMETHYLASE) (LANOSTEROL 14-ALPHA DEMETHYLASE) (LDM) (P450-14DM).
GN CYP51.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxId=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=94296405; PubMed=8024575;
RA Aoyama Y., Funae Y., Noshiro M., Horiuchi T., Yoshida Y.;
RT "Occurrence of a P450 showing high homology to yeast lanosterol 14-
RT demethylase (P450(14DM)) in the rat liver."
RL Biochem. Biophys. Res. Commun. 201:1320-1326(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MISTAR;
RX MEDLINE=98158318; PubMed=9498553;
RA Noshiro M., Aoyama Y., Kawamoto T., Gotoh O., Horiuchi T., Yoshida Y.;
RT "Structural and evolutionary studies on sterol 14-demethylase P450
RT (CYP51), the most conserved P450 monooxygenase: I. Structural analyses
RL of the gene and multiple sizes of mRNA."
RN J. Biochem. 122:1114-1121(1997).
RN [3]
RP SEQUENCE OF 18-503 FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=95394364; PubMed=7665087;
RA Sloane D.L., So O.Y., Leung R., Scarafia L.E., Saidou N., Jarnagin K.,
RA Swinney D.C.;
RT "Cloning and functional expression of the cDNA encoding rat
RT lanosterol 14-alpha demethylase."
RL Gene 161:243-248(1995).
CC -1- FUNCTION: CATALYZES C14-DEMETHYLATION OF LANOSTEROL; IT TRANSFORMS
CC LANOSTEROL INTO 4,4'-DIMETHYL CHOLEST-8,14,24-TRIENE-3-BETA-OL.
CC -1- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) = ROH +
CC OXIDIZED FLAVOPROTEIN + H(2)O.
CC -1- PATHWAY: CHOLESTEROL BIOSYNTHESIS.
CC -1- SUBCELLULAR LOCATION: MICROSMAL (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D55681; BAA09529.1; -
DR EMBL: AB004096; BAA20354.1; -
DR EMBL: AB004087; BAA20354.1; JOINED.
DR EMBL: AB004088; BAA20354.1; JOINED.
DR EMBL: AB004089; BAA20354.1; JOINED.

DR EMBL; AB004090; BAA20354.1; JOINED.
 DR EMBL; AB004091; BAA20354.1; JOINED.
 DR EMBL; AB004092; BAA20354.1; JOINED.
 DR EMBL; AB004093; BAA20354.1; JOINED.
 DR EMBL; AB004094; BAA20354.1; JOINED.
 DR EMBL; AB004095; BAA20354.1; JOINED.
 DR EMBL; U17697; AAA87074.1; -.
 DR InterPro; IPR001128; -.
 DR InterPro; IPR002403; -.
 DR Pfam; PF00067; P450; 2.
 DR PRINTS; PR00465; EP450IV.
 DR PROSITE; PS00086; CYTOCHROME_P450; 1.
 DR OXidoreductase; Monooxygenase; Electron transport; Transmembrane;
 KW Heme; Cholesterol biosynthesis.
 FT TRANSMEM 24 44 POTENTIAL.
 FT BINDING 449 449 HEME (BY SIMILARITY).
 FT CONFLICT 181 181 E -> K (IN REF. 3).
 FT SEQUENCE 503 AA; 56706 MW; 33D8F345FE9CF21 CRC64;

Query Match 13.0%; Score 339; DB 1; Length 503;
 Best Local Similarity 23.7%; Pred. No. 1,4e-16;
 Matches 121; Conservative 97; Mismatches 230; Indels 62; Gaps 16;

QY 6 LLASAL-CTVPLPL-LFLAALKMLDLYCYSSDRSCALPLPGTGM-----PFEFG 55
 DB 26 LLSTLLICAPTLISLVYLFRLAV-----GHVQLPAGANSPPYIYSPITPLG 72
 QY 56 ETLQVAVLQRRKELQKRRKGYFIKTKHLFGRPVVMGADVNRRLILGHLVSVH--WP 113
 DB 73 HAIAGKPIPIELNAYKRYGPVPSFTMGVCTFTYLLGSDAALLFNKNNDLNAEYVG 132
 QY 114 ASVRTILGAGLSNLHDSHKKKKVIMQAFSREALQCYLVIAEYVSSCLEQMLSCGR 173
 DB 133 RLTPPVFGKVAADVPNAVFLQKKILKSGLINAHFKQYVILIEKAEKPYKSWGSEGR 192
 QY 174 GLIYVPEKRLMFRIMLILGCEPAGGEGDEQVLEAEEMRNLT--FS-----LP 225
 DB 193 N--VEAISELITLASHCL-----HGKEIRSQLNEKVAQLYADLDGFSHAAMLTP 242
 QY 226 IDVPFSGLYRGVKA---RNIHARIEENIRAKIRRLQATEPDGCKDALILLIHSMER 281
 DB 243 GMLPLPFSRRRDRRAHREIKNFYKAIQK-----RRL-SKEP---AEDILLIDSTYKD 292
 QY 282 GERLDMQALKQSTSTLLGCGHETTSASATSLITYGLYPLVLOKREIKSGILLCKNQ 341
 DB 293 GRPLTDDDELACMLIGLLAGQHTSSTSAMMGFFLADKKPLQDKCYLDQKT---VCGEDL 349
 QY 342 DNKLDMETLEOLKYIGYIKETLRINPPVPGGFVYALKTFELNGYQIPKGNVIYSIDT 401
 DB 350 P-PLVTEQLKDLNLDRCIKETLRIRLPIMTMRAKTPQYVAGTTPRGHVCVSPVN 408
 QY 402 HDVADIFTKKEEFPNDRFLVPHEDASRFSIPIEGGLSCVCKEPAKILIKITVELAR 461
 DB 409 ORLKDSWVERLDLPNDPRYLQDNPASGEKFAVYFPGAGHRCIGENFAVYQITIMSTMIR 468
 QY 462 HCDMQLNCP-PTMKTSPTVYVVDNLPAEF 490
 DB 469 LYEFDLINGFPPSVNTTMTIPENPVIRY 498

RESULT 10
 CP84_ARATH STANDARD; PRT; 520 AA.
 AC 042600;
 DT 15-DEC-1998 (rel. 37, Created)
 DT 15-DEC-1998 (rel. 37, Last sequence update)
 DT 01-OCT-2000 (rel. 40, Last annotation update)
 DE CYTOCHROME P450 84Y1 (FERULATE-5-HYDROXYLASE) (EC 1.14.-.-) (F5H).
 GN CYP84Y1 OR FAH1 OR AFG36220 OR F23R13.110.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;

OC Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RX MEDLINE=96293440; PubMed=8692910;
 RA Meyer K., Cusumano J.C., Somerville C.R., Chapple C.C.S.;
 RT "Ferulate-5-hydroxylase from Arabidopsis thaliana defines a new
 family of cytochrome P450-dependent monooxygenases.";
 RL Proc. Natl. Acad. Sci. U.S.A. 93:6869-6874(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RX MEDLINE=20083488; PubMed=10617198;
 RA Langham S.-A., McCullagh B., Bilham L., Robben J., Vandebussche F.,
 RA Van der Schueren J., Grymonprez B., Chuang T.-Y., Vandebussche F.,
 RA Braeken M., Weltjens I., Voet M., Bastiaens I., Aert R., Deoor E.,
 RA Weltenegeger T., Bothe G., Ramsperger U., Hilbert H., Braun M.,
 RA Holzer E., Brandt A., Peters S., van Staveren M., Dirkse W.,
 RA Mooljman P., Klein lankhorst R., Rose M., Hauf J., Koetler P.,
 RA Bernerster S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,
 RA De Keyser A., Buysshaert C., Gielen J., Villarroel R., De Clercq R.,
 RA Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,
 RA Clark L., Doggett J., Hall S., Kay M., Lennard N., McClay K., Mayes R.,
 RA Pettitt A., Rajandream M.-A., Lyne M., Benes V., Rechmann S.,
 RA Borkova D., Bloeker H., Scharte M., Grimm M., Loehner T.-H.,
 RA Dose S., de Haan M., Maarse A., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Fartmann B., Granderath K., Danner D., Herzl A.,
 RA Neumann O., Argillou A., Vitale D., Ligouri R., Piravandi E.,
 RA Massenet O., Quidley F., Clabaud G., Muendlein A., Felber R.,
 RA Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,
 RA Cheifor F., Cooke R., Berger C., Monfort A., Casacuberta E.,
 RA Gluhons T., Weber N., Vandenbol M., Barges M., Terol J., Torres A.,
 RA Berbez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,
 RA Heljenz L., Schwarz S., Scholler P., Heber S., Francis P., Bielke C.,
 RA Frishman D., Haase D., Lemcke R.K., de la Bastide M., Habermann K.,
 RA Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,
 RA Parnell L., Dedhia N., Guo J., Schütz K., Hang E., Spiegel L.,
 RA Sebkou M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,
 RA Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,
 RA Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
 RA Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
 RA Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,
 RA Nelson J., Spieth J., Ryan E., Andrews S., Geisel C., Layman D.,
 RA Du H., Ali J., Berghoff A., Jones K., Drove K., Cotton M., Joshi C.,
 RA Antonouli B., Zidanic M., Strong C., Sun H., Lamar B., Jordan C.,
 RA Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shah R.,
 RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Tili S.,
 RA Granat S., Shohdy N., Hasegawa A., Hameed A., Lohli M., Johnson A.,
 RA Chen E., Marra M., Matlienssen R., McCombe W.R.;
 RT "Sequence and analysis of chromosome 4 of the plant Arabidopsis
 thaliana.";
 RL Nature 402:769-777(1999).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. LANDSBERG ERRECTA;
 RX MEDLINE=99097044; PubMed=9880351;
 RA Ruegger M., Meyer K., Cusumano J.C., Chapple C.;
 RT "The regulation of ferulate-5-hydroxylase expression in Arabidopsis in
 the context of sinapate ester biosynthesis.";
 RL Plant Physiol. 119:101-110(1999).
 CC -1- PATHWAY: GENERAL PHENYLPROPANOIC PATHWAY.
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; U88416; AAC49389.1; -

DR EMBL; AL022141; CAA18128.1; -

DR EMBL; AF068574; AD11580.1; -

DR EMBL; AL161589; CAB80293.1; -

DR InterPro: IPR001128; -

DR Pfam: PF00067; P450; 1.

DR PRINTS; PR00385; P450.

DR PROSITE; PS00086; CYTOCHROME_P450; 1.

KW Oxidoreductase; Monooxygenase; Membrane; Heme.

FT BINDING 458 458 HEME (BY SIMILARITY).

SO SEQUENCE 520 AA; 58720 MW; E812779AF5BF01BC CRC64;

Query Match 12.9%; Score 337; DB 1; Length 520;
Best Local Similarity 25.8%; Pred. No. 2,1e-16;
Matches 133; Conservative 90; Mismatches 194; Indels 98; Gaps 23;

OY 15 VLPLLLFLAALIMDLYCVSSRDSCALPLPPTGMPFFEGTLOMYLQRRKFLQMKRK 74
DB 22 VSLPLFTIS-----FTTRRRP---PYPPGPGWPIIGNMMDQLTTRGLANLAKK 70
OY 75 YGFIYKTHL-FGRPTVVMGADNVRRILLGEHRLVSVHVPVGFRTLLGAGCLNLDHSS 132
DB 71 YGGL-CHLRMGFLHMAVSSPEVARQLOVDVSFNRPRT-----AISTLTIDRA 121
OY 133 -----HKORKVIMQAFSREALQCVLV-----IAEVSCLLEQWLSGCGEGL 175
DB 122 DMAFAHYGPFWRQRRKVCVKVFSKRAESVASYRDEVDKVRSSCVNGKPIVNGRQ-- 179
OY 176 LVYPEVKRLMERTAMRLLCCEPAGGDEQOLVFAFEMTR--NLFSIPDIVPPSG 233
DB 180 -ITALRNITTYRAFG--SKCEKG-----QDEFIRILDEFSKLFGAFNVADPIPYFGW 229
OY 234 Y-----RGVKARNLTHARIEENI--RAKIRLQATEPDG-----CKDALQLLEHS 278
DB 230 IDPGIINKRLVAKRNDGFDIDIDEMKKKENQNAVDGVDVDTMDVDDLAFYESEA 289
OY 279 WENGERLDMQ-----ALKOSTELLEGGHETTASATSLITYGLVPHVLOKVEET 330
DB 290 KLVSEPADLQNSIKRLRDNTKAIIMDMVPGTETVASAIMALTELLRSPEDLKRVOEL 349
OY 331 -KSKGLCKSNODNKLDMELEOLKYIGYIKETRLRNPNVPGGFVALKTFELNGYQIP 389
DB 350 AEVYGL-----DRVDESIEKILTYIKTLEKTRMHPPIPLLHETADTSDIGFFIP 403
OY 390 KGMNVYISICDTHDVADIFTNKEEFPNDRIIVPHED--ASRFSFIPFGGLSCVCKEF 447
DB 404 KKSRYVINAFALGRDPTSMWDPDTPFRSRLERGVDPFKSNFEIPIFGSGRSCPPOMQ- 462
OY 448 AKLLKIFVELA-----RHC-DWQLLNGPPTMTS 477
DB 463 ----LGLVALDLAVAHILHCTWKL--PDGMPKS 490

RESULT 11
CP51_PIG STANDARD; PRT; 503 AA.

AC 046420;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE CYTOCHROME P450 51 (EC 1.14.14.1) (CYPL1) (P450L1) (STEROL 14-ALPHA DEMETHYLASE) (LANOSTEROL 14-ALPHA DEMETHYLASE) (LDM) (P450-14DM).
GN CP51.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suidae; Suidae; Sus.

OX NCBI_TaxID=9623;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Kojima M., Morozumi T., Hamasima N., Okamoto T.;
RT "Cloning of a pig lanosterol 14-demethylase cDNA";
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: CATALYZES C14-DEMETHYLATION OF LANOSTEROL. IT TRANSFORMS LANOSTEROL INTO 4,4'-DIMETHYL CHOLESTA-8,14,24-TRIENE-3-BETA-OL (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) = ROH + OXIDIZED FLAVOPROTEIN + H(2)O.
CC -1- PATHWAY: CHOLESTEROL BIOSYNTHESIS.
CC -1- SUBCELLULAR LOCATION: MICROSOMAL (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; AB009988; BAA24134.1; -

DR InterPro: IPR001128; -

DR InterPro: IPR002403; -

DR Pfam: PF00067; P450; 2.

DR PRINTS; PR00465; EP450TV.

DR PROSITE; PS00086; CYTOCHROME_P450; 1.

KW Oxidoreductase; Monooxygenase; Electron transport; Transmembrane;
KW Heme; Cholesterol biosynthesis.
FT TRANSMEM 24 44 POTENTIAL.
FT BINDING 449 449 HEME (BY SIMILARITY).
SO SEQUENCE 503 AA; 56866 MW; 0302949CE461AED6 CRC64;

Query Match 12.7%; Score 332; DB 1; Length 503;
Best Local Similarity 23.2%; Pred. No. 4.4e-16;
Matches 117; Conservative 97; Mismatches 239; Indels 52; Gaps 14;

OY 6 LLASAL--CTFVPLLLFLAALIMDLYCVSSRDSCALPLPPTG-----FPFGE 56
DB 26 LLSSLLACAFLL-ILVYLFQALGHL-----APLPAGAKSPYIFSPILPLGH 73
OY 57 TLQVNLQRRKFLQMKRKXGEIYKTHL-FGRPTVVMGADNVRRILLGEHRLVSVH--WPA 114
DB 74 ALAFGKSPITFLENAYEKYGPVFSFTVWGFYLLGSDAALLFNSKNEDLNAEDVYSR 133
OY 115 SVRTILGAGCLSNLHDSHQKRVIMQAFSREALQCVLVIAEVSCLLEQWLSGCGE 174
DB 134 LTPPVFGKVAAYVNPVNFLEQKKMLKSGLINAHFRHVSIIIEKTEYQSGESERN 193
OY 175 LLYPEVKRLMFRNARILLGCEPAGGDEQOLVFAFEMTRNL--FS-----LPT 226
DB 194 L--FEALSELIIITLASHCL-----HGKEIRSOILNEKVAQLYADLDGFGFSAAMLPLG 243
OY 227 DVPESGLRGVKARNLTHARIEENINAKIRLQATEPDGCKALQILLHSMERGRLD 286
DB 244 WLPSPFRR-----RDRAHREIKNIIFYKAIQKROSEK--IDILQTLDTSTYKDGPRLT 297
OY 287 MQALKOSTELLEGGHETTASATSLITYGLVPHVLOKVEIKSGGLCKSNODNKLD 346
DB 298 DDEVAGMLIGLLAGQHTSTTSAMMGFPLARKTLOEKYLOKTI--VCGSDLP-PLT 353
OY 347 MELEOLKYIGYIKETRLRNPNVPGGFVALKTFELNGYQIPKGMNVYISICDTHDVAD 406
DB 354 YDQLKDLNLDIRKIKETRLRPNIMTMRMAKTPQYVAGTTPGHQVCSPTVNORLKK 413
OY 407 IFNKEEFPNDRIIVPHEDASRFSFIPGCGLRSCVNGKFEAKILKIFVELARRCDMG 466
DB 414 SWVERLDENPDRILQNPASGEFAFYVPGAGNHCRTIGENFAVQIKTITWSTMLRYEFD 473

QY 467 LINGP-PTKSTPTVYVNDLPARF 490
DB 474 LIDGEPVNYTMTHTPENPYRY 498

RESULT 12
C4D8_DROME STANDARD; PRT; 505 AA.

AC 09V579; 024127;
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE CYTOCHROME P450 4D8 (EC 1.14.-.-) (CYP11B8).
GN CYP4D8 OR CG4321.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Cealiker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Milos G.L.G.,
RA April J.F., Aghayani A., An H.-J., Andrews-Pfankuch C., Baldwin D.,
RA Ballou R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Bottier P.,
RA Burtils K.C., Busam D.A., Butler H., Cadiou E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbalt W.M., Glaser K.,
RA Glodde A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegawa C.,
RA Jallat M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Klotz C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Mosnig A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirasas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao O.A.,
RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Glibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster";
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE OF 315-442 FROM N.A.
RC STRAIN=HAAG-79;
RX MEDLINE=96262181; PubMed=8676871;
RA Dunkov B.C., Rodriguez-Arnal R., Pittendrigh B.,
RA French-Constant R.H., Feyereisen R.;
RT "Cytochrome P450 gene clusters in Drosophila melanogaster";
RL Mol. Genet. 251:290-297(1996).
CC -1- FUNCTION: MAY BE INVOLVED IN THE METABOLISM OF INSECT HORMONES AND
CC IN THE BREAKDOWN OF SYNTHETIC INSECTICIDES (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) -> ROH +
CC OXIDIZED FLAVOPROTEIN + H(2)O.

CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND, ENDOPLASMIC RETICULUM
CC (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.1sb-sib.ch/announce/>
CC or send an email to license@1sb-sib.ch).
CC -----
CC EMBL: AE003558; AAF50549.1; -
DR EMBL: U34329; AAA80663.1; -
DR HSSP: P14779; 1B07.
DR Flybase: FBgn0015033; Cyp4d8.
DR InterPro: IPR001128; -
DR Pfam: PF00067; P450; 1.
DR PRINTS: PS00385; P450.
DR PROSITE: PS00086; CYTOCHROME_P450; 1.
KW Oxidoreductase; Monooxygenase; Membrane; Heme; Microsome;
KW Endoplasmic reticulum.
FT BINDING 451 451 HEME (BY SIMILARITY).
FT CONFLICT 333 333 A -> E (IN REF. 2).
FT CONFLICT 435 435 R -> C (IN REF. 2).
SQ SEQUENCE 505 AA; 57819 MW; 7BA5271ED46093F CRC64;

Query Match 12.5%; Score 327.5; DB 1; Length 505;
Best Local Similarity 26.5%; Pred. No. 9.2e-16;
Matches 131; Conservative 87; Mismatches 215; Indels 61; Gaps 18;

QY 14 FVLPLLLFLAALAKMDLYCVSSRSCALPLPGTGMFPFEE-TLQWLYORRFLQMKR 72
DB 4 FLLVLLLFAG---WITH-LGQADRKRKANIPGLCPPLIGAMQMLRLNKTIYKGR 59
QY 73 R--KXGYFTKHLFGPPTVRYVGDVNRILLGHRVSVHPASVRTIGAGLSNLH 129
DB 60 EYLRKFGHQRWIFENRLLIMSGDAELNQLSSQHLVK---HPVYKVLQWMLNGILL 115
QY 130 DSKHK---QRKVIQAFAFREALQCVLYIAEVSCLQWMLSCGRGLLYPEVKRLMF 186
DB 116 LSGKXWHQRKRITPTFHSILDEQVEVEFDQSNICVQRLAQANGTF---DYRSIC 172
QY 187 RIMRILLGCEPRA--GGGEQQLVEAFEEKTRNL-----FSLPDPVPSGLYRGK 238
DB 173 AALDIITAEIAGMTKTYAQAANESTPFAEAVNETALSRFMSVYIQVLLFTLHPHLK 232
QY 239 AR--NLIHARIEENIRAKIRLQATE-----PDGCKDALQL--IEHSMERGE 283
DB 233 WRQTLIRFMQEFITIKYIEKRQALDEQSKLMDTADDEVGSKRRRALLDVLMSTVDGR 292
QY 284 RLDMQALKOSTELLGGHTTASATSLITLYGLPHYLQVREIRKSKGLICKSNQDN 343
DB 293 PLNDELREVDVTFMEGHDTTALSFCHELSRHPVQAKLEIYO---VLGIDRSR 349
QY 344 KLDMETLEOKYIGYIKETRLNPPVPGGFRVALKTFEL-----NGYQIDPGMNVIY 397
DB 350 PVSIRLDGLKIVKCYIKESLRMPVPIYGRKLQIDFKYTHSVHGDV-IPAGSIIIG 408
QY 398 ICDTHADVAFITNKEFPNDRFTVPHEDASR---FSFIPGGGLSCYCKGEKAKILKI 454
DB 409 IFGVHROPFTFPNPDDEFIPRH-----ENGSRAVAPFKMIPFSAGPRNIGOKFAOLEMK 463
QY 455 FYVEIARHCQMOIL 468
DB 464 MLAKIYR--EYELL 475

RESULT 13
C132_DROME STANDARD; PRT; 493 AA.
AC 09V644;

DT 01-OCT-2000 (Rel. 40, Created)
 DT 01-OCT-2000 (Rel. 40, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE PROBABLE CYTOCHROME P450 313A2 (EC 1.14.14.1) (CYPCXXIIA2).
 GN CYP313A2 OR CG10094.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 NCBI_Taxid=7227;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BERKELEY.
 RX MEDLINE-20196006; PubMed-10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer V.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfankuch C., Baldwin D.,
 RA Ballow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borova D., Botchan M.R., Bouck J., Brockstein P., Brotlier P.,
 RA Burks K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durkin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Fostel C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman J.D., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeigbam C.,
 RA Jaitli D., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mottel B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle V., Reese M.G.,
 RA Palazuelo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Relbert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kimlos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 "The genome sequence of Drosophila melanogaster";
 RL Science 287:2185-2195(2000).
 CC -1- FUNCTION: MAY BE INVOLVED IN THE METABOLISM OF INSECT HORMONES AND
 CC IN THE BREAKDOWN OF SYNTHETIC INSECTICIDES (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) -> ROH +
 CC OXIDIZED FLAVOPROTEIN + H(2)O.
 CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM
 CC (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib.ch).
 CC
 CC EMBL, AE003695; AAE54769.2;
 DR FLYBase; FBgn0038006; Cyp313a2.
 DR InterPro; IPR001128;
 DR Pfam; PF00067; P450; 1.
 DR PRINTS; PR00385; P450.

DR PROSITE; PS00086; CYTOCHROME_P450; 1.
 KW Oxidoreductase; Monooxygenase; Membrane; Heme; Microsome;
 KW Endoplasmic reticulum; Hypothetical protein.
 FT BINDING 438 438 HEME (BY SIMILARITY).
 SQ SEQUENCE 493 AA; 56651 MW; 808769FAD0627262 CRC64;
 Query Match 12.5%; Score 326.5; DB 1; Length 493;
 Best Local Similarity 25.2%; Pred. No. 1e-15;
 Matches 122; Conservative 88; Mismatches 238; Indels 37; Gaps 14;
 QY 19 LFLPALKIMDIVCVSRDSCALPLPPGTMGPPFFBETIQMYLQRRKFLQMKR---KY 75
 DB 6 LTLAASLILMIRFLSRRKRLKMLMQLPGRMRGLPLGNSVYLLISGRSSFTTYMDKH 65
 QY 76 GFLYKTHLFGRPYRVAGDANVRRIILGEHRL-VSWMPASVFTILGAGCLSNLHDSHK 134
 DB 66 GSTYMAWIGTTPYITRDPIAEKVLSPSPICINRSSQTTNALMSMGLLT-LQSKMM 124
 QY 135 QRRKVMQARSREALOCYVIVAEVS---SLEQWLSCERGLVPEYKRLMFRIMR 191
 DB 125 ARRRHMPARFHSVLLSELPFMAETDLVSPDSFVGQGEKVL--SDLIRMSFALATQ 182
 QY 192 ILGCEPFGAGGEDDEQOLVAFEEEMTRNLSLPIDVF-----SGLYGVKARNLIHA 245
 DB 183 TTIGTDTKDNFEND-AILKTYQSMKR-LTIINIFPYPQNKIVSKLFGLEMLRRDAS 240
 QY 246 RIEENIRAKRRLQAEVPPDGCKDAQLILHSWE--RGERLDMALQKOSTELLEFGHE 303
 DB 241 AIKMINNIDKKLINSPEVCESELTQYIHRALIEFRNDEMELMELGACSSMVAFAE 300
 QY 304 TTASATSLITLYGLVPHYLQVREIFKSKGLICKSNQDKKLMETLEQKLYGCIYKER 363
 DB 301 TSHATYVALVLLAMPPEHDEMFEIKELFPLAKGLEVTHTD--LQQLVYDRVLNET 357
 QY 364 LRLNPPVGGFRRVALKTFEL-NGQYLPKGNVYISICDTHVDADIF-TNKEFNPDRFI 421
 DB 358 LRLMPGVPSRRFLDLRLNSGVVIPKGMTISIDFNGRNDVNGSEAGQNPENFLP 417
 QY 422 PHEDDSRSFSTFPPGGGLSCVCEKFAKILKFTVELARHCDMOLLNPPMTKTSPTY 481
 DB 418 EKTHDHPPAFPIFSPKGRKRCIGMRYLGLMSSKLALVKILRNY-----KLKTS--F 465
 QY 482 PVDNL 486
 DB 466 PYENL 470
 RESULT 14
 C4D2_DROME STANDARD; PRT: 501 AA.
 AC 027589; 027588; 046053; 018651; 018674; 09W514;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 01-OCT-2000 (Rel. 40, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE CYTOCHROME P450 4D2 (EC 1.14.14.1) (CYP4D2).
 GN CYP4D2 OR EG:152A3.4 OR CG3466.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Ephydroidea; Drosophilidae; Drosophila.
 NCBI_Taxid=7227;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-OREGON-R;
 RX MEDLINE-94296569; PubMed-8024706;
 RA Fedlov M.V., Alatorsev V.E.;
 RT "Cluster of cytochrome P450 genes on the X chromosome of Drosophila
 RL DNA cell Biol. 13:663-668(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-OREGON-R;

RX MEDLINE=20196011; PubMed=10731137;
 RA Benos P.V., Gatt M.K., Ashburner M., Murphy L., Harris D.,
 RA Barrett B.G., Ferraz C., Vidal S., Brun C., Demillies J., Cadieu E.,
 RA Dreno S., Gloux S., Lelaure V., Mottier S., Galibert F., Borkova S.,
 RA Minana B., Katios E.C., Louis C., Siden-Klamos I., Bolshakov S.,
 RA Papadogiannakis G., Spanos L., Cox S., Madeno E., de Pablo B.,
 RA Madoiell J., Peter A., Schottler P., Werner M., Mouriklotti F.,
 RA Beinert N., Dove G., Schafer U., Jackle H., Bucheton A.,
 RA Callister D.M., Campbell L.A., Darlamsitso A., Henderson N.S.,
 RA McMillan P.J., Salles C., Tait E.A., Valenti P., Saunderson R.D.,
 RA Glover D.M.,
 RT "From sequence to chromosome: the tip of the x chromosome of D.
 RT melanogaster.";
 RL Science 287:2220-2222(2000).
 RN [31]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amentides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Vandal M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Bernier B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brockstein P., Brotler P.,
 RA Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delecher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Dushin K.J., Evangelista C.C., Ferraz C., Fertiera S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glasser K.,
 RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jaitai M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P.E., Leal Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy N., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spletter E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Spierker R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissensbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 RN [4]
 RP SEQUENCE OF 4-501 FROM N.A.
 RC STRAIN-VARIOUS STRAINS:
 RA Phillips R.S., Begun D.J., Aquadro C.F.;
 RT "Evidence for non-neutral evolution around the cytochrome p450 gene
 RT cluster on the drosophila melanogaster x chromosome.";
 RL Submitted (OCT-1997) to the EMBL/Genbank/DBJ databases.
 CC -1- FUNCTION: INVOLVED IN THE METABOLISM OF INSECT HORMONES AND IN
 CC THE BREAKDOWN OF SYNTHETIC INSECTICIDES (BY SIMILARITY)
 CC -1- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) = ROH +
 CC OXIDIZED FLAVOPROTEIN + H(2)O.
 CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM
 CC (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

SEQ	SEQUENCE	501 AA:	57914 MW:	8FBD98DBF39FD1A2 CRC64:
DR	EMBL; X/35955; CAAB3508.1; -			
DR	EMBL; Z23005; CAAB0549.1; -			
DR	EMBL; AL009194; CAAF5698.1; -			
DR	EMBL; AE003423; CAA45741.1; -			
DR	EMBL; AF017006; AAB71169.1; -			
DR	EMBL; AF017007; AAB71170.1; -			
DR	EMBL; AF017008; AAB71171.1; -			
DR	EMBL; AF017009; AAB71172.1; -			
DR	EMBL; AF017010; AAB71173.1; -			
DR	EMBL; AF017011; AAB71174.1; -			
DR	EMBL; AF017012; AAB71175.1; -			
DR	EMBL; AF017013; AAB71176.1; -			
DR	EMBL; AF017014; AAB71177.1; -			
DR	EMBL; AF017015; AAB71178.1; -			
DR	EMBL; AF017016; AAB71179.1; -			
DR	EMBL; AF017017; AAB71180.1; -			
DR	EMBL; AF017018; AAB71181.1; -			
DR	FLYBASE: FBgn0011576; Cyp4g2.			
DR	InterPro: IPR001128; -			
DR	InterPro: IPR002402; -			
DR	Pfam: PF00067; P450; 1.			
DR	PRINTS: PRO0385; P450.			
DR	PRINTS: PRO0464; EP450II.			
DR	Prosite: PS00086; CYTOCHROME_P450; 1.			
KM	Oxidoreductase: Monooxygenase; Membrane; Heme; Microsome;			
KM	Endoplasmic reticulum; Polymorphism.			
FT	BINDING	449	449	
FT	VARIANT	163	163	
FT				HEME (BY SIMILARITY).
FT				K -> M (IN STRAINS CAM-44, CAM-48 AND
FT				BERKELEY).
FT	CONFLICT	2	25	MISSING (IN REF. 2).
FT	CONFLICT	30	30	I -> A (IN CAAB0549).
FT	CONFLICT	160	160	A -> R (IN REF. 1).
FT	CONFLICT	487	501	LSANGVHGGLKPR -> CGRPATFILA (IN REF.
FT				1).

[illegible]

Db 354 TLRLDGLKEMENYIKESLRHPPVIMGRWAEDEVRGKHIPAGTMTGIFVLLRDP 413
 QY 406 DIFNKEEFNDRE--IYP--HPEDASRSPFLPGGGRSCVCKEFAILKIFVELAR 461
 Db 414 EYFSPDEFEREDADVPQIHP-----YAIIPISAGRNCIGOKFAMLEKSTVSKLLR 468
 QY 462 HCDWOLNNGPP 472
 Db 469 HFELPLGPEP 479
 RESULT 15
 CP51_HUMAN STANDARD; PRT: 503 AA.
 ID CP51_HUMAN 016850; 016784; 000770; 000772; Q99868;
 AC 01-NOV-1997 (Rel. 35, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE CYTOCHROME P450 51 (EC 1.14.14.1) (CYP11) (P450L1) (STEROL 14-ALPHA
 DEMETHYLASE) (LANOSTEROL 14-ALPHA DEMETHYLASE) (LDM) (P450-14DM).
 GN CYP51.
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN 11
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 MEDLINE=96201125; PubMed=8619637;
 RA Stroemstedt M., Rozman D., Waterman M.R.;
 RT "The ubiquitously expressed human CYP51 encodes lanosterol 14 alpha-
 demethylase, a cytochrome P450 whose expression is regulated by
 RT oysterols.";
 RL Arch. Biochem. Biophys. 329:73-81(1996).
 RN 12
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 Noshiro M.;
 RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
 RN 13
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 MEDLINE=96404948; PubMed=8809088;
 RA Rozman D., Stroemstedt M., Waterman M.R.;
 RT "The three human cytochrome P450 lanosterol 14 alpha-demethylase
 (CYP51) genes reside on chromosomes 3, 7, and 13: structure of the
 RT two retrotransposed pseudogenes, association with a line-1 element,
 RT and evolution of the human CYP51 family.";
 RL Arch. Biochem. Biophys. 333:466-474(1996).
 RN 14
 RP SEQUENCE FROM N.A.
 MEDLINE=97131516; PubMed=8975714;
 RA Rozman D., Stroemstedt M., Tsui L.-C., Scherer S.W., Waterman M.R.;
 RT "Structure and mapping of the human lanosterol 14alpha-demethylase
 RT gene (CYP51) encoding the cytochrome P450 involved in cholesterol
 RT biosynthesis: comparison of exon/intron organization with other
 RT mammalian and fungal CYP genes.";
 RL Genomics 38:371-381(1996).
 CC -1- FUNCTION: CATALYZES C14-DEMETHYLATION OF LANOSTEROL; IT TRANSFORMS
 CC LANOSTEROL INTO 4,4'-DIMETHYL CHOLESTRA-8,14,24-TRIENE-3-BETA-OL.
 CC -1- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) = ROH +
 CC OXIDIZED FLAVOPROTEIN + H(2)O.
 CC -1- PATHWAY: CHOLESTEROL BIOSYNTHESIS.
 CC -1- SUBCELLULAR LOCATION: MICROSOAL (POTENTIAL).
 CC -1- TISSUE SPECIFICITY: UBQUITOUSLY EXPRESSED WITH HIGHEST LEVELS IN
 CC TESTIS, OVARY, ADRENAL, PROSTATE, LIVER, KIDNEY, AND LUNG.
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U23942; AAB39951.1; ALT_INIT.
 DR EMBL; D55653; BAA09512.1; ALT_INIT.
 DR EMBL; AC000120; AAB46356.1; ALT_INIT.
 DR EMBL; U51692; AAC50951.1; -.
 DR EMBL; U51684; AAC50951.1; JOINED.
 DR EMBL; U51685; AAC50951.1; JOINED.
 DR EMBL; U51686; AAC50951.1; JOINED.
 DR EMBL; U51687; AAC50951.1; JOINED.
 DR EMBL; U51688; AAC50951.1; JOINED.
 DR EMBL; U51689; AAC50951.1; JOINED.
 DR EMBL; U51690; AAC50951.1; JOINED.
 DR EMBL; U51691; AAC50951.1; JOINED.
 DR MIM; 601637; -.
 DR InterPro: IPR001128; -.
 DR InterPro: IPR002403; -.
 DR Pfam: PF00067; P450; 2.
 DR PRINTS; PR00465; EP4501V.
 DR PROSITE; PS00086; CYTOCHROME_P450; 1.
 KW Oxidoreductase; Monooxygenase; Electron transport; Transmembrane;
 KW Heme; Cholesterol biosynthesis.
 FT TRANSMEM 24 44
 FT BINDING 449 449 HEME (BY SIMILARITY).
 FT CONFLICT 272 272 R -> T (IN REF. 4).
 FT CONFLICT 368 368 K -> R (IN REF. 4).
 SQ SEQUENCE 503 AA; 56805 MW; 4FCBC147B4DEDB6 CRC64;

Query Match 12.5%; Score 326; DB 1; Length 503;
 Best Local Similarity 23.5%; Pred. No. 1,2e-15;
 Matches 119; Conservative 94; Mismatches 240; Indels 54; Gaps 13;
 QY 3 LPALLASALCTEVALPLLELAALKLMDLYCVSSRDRSCALPLPGTNG-----PPFG 55
 Db 27 LSMILACAFLLSLVYLIRLAGHL-----VOLPAGVKSPPYIFSPIFLG 72
 QY 56 ETLQWLVQRKFLQMRKRGKGLTKTLPGRPYRVGADNVRKILLGRLSVH--WP 113
 Db 73 HATAFGKSPTEFLENAYEKGVFSTWVGKTFYLLGSDAALLFNSKMDLNAEDVYS 132
 QY 114 ASVYTIAGACLSNLMDSSHQRKRYLQAFSREALQCYLVIAEYVSCLEQWLSCGR 173
 Db 133 RLTPYVGVKAVDVPNPVLEQKKMLKSLNIAHFRQHSIIETKEKKEFEESGEGEK 192
 QY 174 GLIYVEYKRLMRIRAMRILLGCEPAGGEGEDQULVEAFEEKTNL---FS----LP 225
 Db 193 N--VFALSELILITASHHC-----HKEIRSOINEKVAOLYADLDGFSHAAMLIP 242
 QY 226 IDVPEGLYGVKARNLIHRIEENI-RAKIRRLQATEPDGGCKDLQLLIESWGER 284
 Db 243 GWPLPLSFRR-----RRAHREIDITYKALQKRQOEK---IDDLIQLLDDTYDGR 295
 QY 285 LDMQALKOSTELLFSGHETTAAATSLITYGLVHYLVQKVEEIKSKGLCKSNQDNK 344
 Db 296 LTPDEVAGMLIGLLAGOHSTSTTSAMGFFLARDKTLQKKCYLEQKT--VCGENLP-P 351
 QY 345 LDMETPEOLKIGCYIKETRLNPPVGGFRVALKFFELNGYIIPGMANYTISICTHDV 404
 Db 352 LTYDQKLDNLDRCKIKETRLRPTIMIMRMARPDQVAGYITPGHOCVSPYNQRL 411
 QY 405 ADIFNKEEFNDREITVPEDASRSPFLPGGGRSCVCKEFAILKIFVELARHCD 464
 Db 412 KDSWVERLDENPRYLDQDPASGEKFAVYFPGAGRRCIGENAVYQITWSTMRLRYE 471
 QY 465 NQULNRP-PTMKTSPYIVYDNLNPARF 490
 Db 472 FDLIDYFPYTVNTYIMIRHPENPVIRY 498


```

107 GCTGCTCTTCTTCCGCGCGCTCAAGCTCTGGAGACTCTGACTGTGTGA 156
34  eSerArgAspArgSerCysAlaLeuProLeuProProGlyThrMetGly 50
157 GCAAGCGGACATCGACAGTGGCCCTCCCTGGCCCCGGGTACCATGGGC 206
51  pheProPhePheGlyLysThrLeuGlnMetValLeuGlnArgArgLysPhe 67
207 TTCCCATCTTTGGGGGAACATTGCAGATGTGCTTCACGGGAGGAAGTT 256
67  eleuGlnMetLysArgArgLysThrGlyPheIleTyrLysThrHisLeuP 84
257 TCTGCAGATGAAGCCGACGAAATACGGCTTCATCTCAAGACCATCTGT 306
84  heGlyArgProThrValArgValMetGlyAlaAspAsnValArgArgGly 100
307 TTGGCGGCGCCACAGGTGCGGTGATGGCGCGGATATATGTGGCGGCATC 356
101 LeuLeuGlyGlnHisArgLeuValSerValHisTrpProAlaSerValAr 117
357 TTGCTGGGAGACACCGGTTGGTGGTGCACTGGCCCGCGTGGTGG 406
117 gThrIleLeuGlyAlaGlyCysLeuSerAsnLeuHisAspSerSerHisL 134
407 CACCATCTGGGCGGTGGCTGCTCCAACTGCACGATTCCTCGCACAA 456
134  ysGlnArgLysLysValIleMetGlnAlaPheSerArgLysAlaLeuGln 150
457 ACCAGCGAAGAAAGGTGATTATGACAGGCTTCAGCGCGGACACCTCCAG 506
151 CysTyrValLeuValIleAlaGlnGluValSerSerCysLeuGlnGlnTr 167
507 TTCTACGTGCTGCTGATCGCTGAGGAAGTCACAGATTGCTGAGACAGTG 556
167 pleuSerCysGlyGluArgGlyLeuLeuValTyrProGluValLysArgL 184
557 GCTAAGCTGCGCGGAGCGGCGCTCTGCTGTACCCCGAGGTGAAGCGCC 606
184 eumetPheArgIleAlaMetArgIleLeuLeuGlyCysGlnProGlyPro 200
607 TCATGTTCGCGATCGCCATGCGCATCTGCTGGGTGGAGCGGGGTCCA 656
201 AlaGlyGlyGlyGluAspGlnGlnGlnLeuValGlnAlaPheGlnGluMe 217
657 GCGGGCGGCGGGAGGAGCAGCACAGCTCGTGGAGGCTTTGAGAGAGAT 706
217 tThrArgAsnLeuPheSerLeuProIleAspValProPheSerGlyLeuT 234
707 GACCGGCAATCTCTCTCTCTCCCATTTGACGTGCCCTTTAGCGGCTGT 756
234 yTrArgGlyValLysAlaArgAsnLeuIleHisAlaArgTleGlnGluAsn 250
757 ACCGGGGGTGAAGGCGCGAATTAACAGCGGCATCGAGAGGAAC 806
251 IleArgAlaLysIleArgArgLeuGlnAlaThrGlnProAspGlyGlyCy 267
807 ATTCCGCGCAAGATCCCGGCTTACGCTTACAGAGCGGAGTGGGGTGG 856
267 sLysAspAlaLeuGlnLeuLeuIleGlnHisSerTrpGluArgGlyGluA 284
857 CAAGGACGCGCTGCAGTCTGATTGAGCACCTCGTGGAGAGGGAGAGA 906
284 rGluLeuAspMetGlnAlaLeuLysGlnSerSerThrGluLeuLeuPheGly 300
907 GGGCTGGATATCGAGCACTAAACATGTCACACAGACCTCTCTTTGGT 956
301 GlyHisGluThrThrAlaSerAlaIleThrSerLeuIleThrTyrLeuG 317
957 GGTATGAACATTACAGCCAGTGTGCGACATCAGTACACTTACTTACCTAGG 1006
317 yLeuTyrProHisValLeuGlnLysValArgGlnGluIleLysSerLysG 334
1007 ACTTACCCACATGTCTCCGAGAAAGTTGAGAAAGATTAAGAGCAAGG 1056

```

```

334  LysLeuCysLysSerAsnGlnAspAsnLysLeuAspMetGluThrLeu 350
1057 GCTTACTTTGGCAAGACAAATCAAGACAAACAGTTAGACATGAAACTTGG 1106
351  GluGlnLeuLysTyrIleGlyCysValIleLeuGlnThrLeuArgLys 367
1107 GAACAGCTTAATATCATTTGGTGTGTCAATTAAAGAGACCTCGGATTGAA 1156
367  nProProValProGlyGlyPheArgValAlaLeuLysThrPheGluLeuA 384
1157 TCTCCGGTCCCGAGAGGGGTTTCGGGTGCTCGAAGACTTTGACTGTA 1206
384  sngLysTyrGlnIleProLysGlyTrpAsnValIleTyrSerIleCysAsp 400
1207 ATGGATACAGATCCCAAGGGCTGGAATGTTATTATACATATCTGTGAC 1256
401  ThrHisAspValAlaAspIlePheThrAsnLysGlnGluPheAsnProAs 417
1257 ACCACAGATGTGGCAGATATCTTCACTAACAGAGAGAAATTATCCCA 1306
417  PArgPheIleValProHisProGluAspAlaSerArgPheSerPheIleP 434
1307 CCGCTTATATAGTCCCTCATCCAGAGGATGCTCCCGTTACAGCTTCATTC 1356
434  ropheGlyGlyGlyLeuArgSerCysValGlyLysGlnPheAlaLysIle 450
1357 CATTTGAGAGAGGCTTCGAGCTGTGTAGCAAAAGATTGCAAAATTT 1406
451  LeuLeuLysIlePheThrValGlnLeuAlaArgHisCysAspTrpGlnLe 467
1407 CTTCTTAAGATATTATACAGTGAGCTGGCTAGGCACTGTGATTGGCAGCT 1456
467  uleuAsnGlyProProThrMetLysThrSerProThrValTyrProValA 484
1457 TCTAAATGACCTCTCAATAGAACAAAGCCCACTGTGTACCTGTGG 1506
484  sPAsnLeuProAlaArgPheThrTyrPheGlnGlyAspIle 497
1507 ACAATCTCCCTGCACATTCACCTACTTCCAGAGAGATATC 1547

seq_name: gb_f01:AF115769
seq_documentation_block:
LOCUS      AF115769      1726 bp      mRNA      03-MAR-1999
DEFINITION Mus musculus cytochrome P450 retinoic acid metabolizing enzyme
ACCESSION AF115769
VERSION    AF115769.1 GI:4324987
KEYWORDS   house mouse.
SOURCE     Mus musculus.
ORGANISM   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 1726)
AUTHORS   Paine,C.T., Paine,M.L. and Sneed,M.L.
TITLE      Identification of amelogenin- and tuftelin-interacting proteins
using the yeast two-hybrid system
Connect. Tissue Res. 38, 257-267 (1998)
JOURNAL    2 (bases 1 to 1726)
Paine,C.T., Paine,M.L. and Sneed,M.L.
AUTHORS    Direct Submission
TITLE      Submitted (21-DEC-1998) CCMB, University of Southern California,
JOURNAL    2250 Alcazar Street, CSA103, Los Angeles, CA 90033, USA
FEATURES   location/Qualifiers
source     1..1726
            /organism="Mus musculus"
            /db_xref="taxon:10090"
            /tissue_type="tooth"
            /dev_stage="1 day old"
            27..1520
            /note="tuftelin-interacting protein 24 TIP-24"
            /codon_start=1

```

```
/product="cytochrome P450 retinoic acid metabolizing
enzyme P450RA"
/protein_id="AAD17217.1"
/db_xref="GI:4324988"
/translation="MGIPALLATATCTFVLPILLFLAALKIMDLVCSSRDRSCALPL
PGTGMFPFPGFTLQMTVLQRRKFLQMRKRKRYFYTKTLFGSPYRWGADNVRILL
GHRIVSVHWPASVPRITLGAGLSNLHDSHKQRKRLMQASREALOCYUVAEEV
SSCLQWLSCGERGLVPEVRKLMERIAMILLCEPDPAGGSEDOOLVEAFEMT
RILFSLPIDVPFGSLRGVKARNLIHARIEENIRAKIRKLQATEPDGCKDALQLIE
HSMERGERLDMOLKOSSTELFGHETASATSLIYLGLYPHYLOKVRREIKSG
LCKSNODMKLMELEOLKTYGVYIKETRLNPNVPGFVALTFELNGYQIPKGM
NWTISCDPHVDADIPTNKEEPNPRFTVPHEDASRSFPIPEGGSLSCVKEPAKI
LKITFVELARHCDMLNGPPMTSPVIVFVDNLPAFTTFQGDII"
BASE COUNT      405 a      428 c      476 g      417 t
ORIGIN
alignment_scores:
  Quality: 2609.00      Length: 497
  Ratio: 5.249      Gaps: 0
  Percent Similarity: 100.000      Percent Identity: 99.799
alignment_block:
US-09-668-482-32 x AF115769 ..
Align seg 1/1 to: AF115769 from: 1 to: 1726
1 MetGlyLeuProAlaLeuLeuAlaSerAlaLeuCysThrPheValLeuPr 17
|||||
27 ATGGGGCTCCCGCGCTGCTGCGCACCTGCGCTGACCTTCGTCGCC 76
|||||
17 OleuleuLeuPheLeuAlaAlaLeuLysLeuTrpAspLeuTyrCysValS 34
|||||
77 GCTGCTGCTCTTCCTGCGGCGCGCTCAAGCTCGGACCTGACTGTGTA 126
|||||
34 enSerArgPargSerCysAlaLeuProLeuProGlyThrMetGly 50
|||||
127 GAGCGCGCATGCGACGTGCGCCCTCCCTCCCGGTCACCTATG 176
|||||
51 PheProPhePheGlyLysThrLeuGlnMetValLeuGlnArgArgLysPh 67
|||||
177 TTCCTATCTTTGGGGAACATTGCAGATGTGCTTCAGCGGAGAGATT 226
|||||
67 eleuGlnMetLysArgArgLysTyrGlyPheLeuTyrLysThrHisLeuP 84
|||||
227 TCTGCAGATGAAGCGCAGGAAATAGCGCTTCATCAAGACGACATCTGT 276
|||||
84 heGlyArgProThrValArgValMetGlyAlaAspAsnValArgArgLys 100
|||||
277 TTGGGGCGGCCACGGGTGCGGTGATGGCGCGGATATGTGGCGGCATC 326
|||||
101 LeuLeuGlyLysHisArgLeuValSerValHisTrpProAlaSerValAr 117
|||||
327 TTCCTGGAGAGACCGGTGTGCTGCTGCACTGCGCCGCTCGGTGCG 376
|||||
117 gThrIleLeuGlyAlaGlyCysLeuSerAsnLeuHisAspSerSerHisL 134
|||||
377 CACCATCTCGGGCGCTGCTGCTCCAACTGCACGATTCCTCCACA 426
|||||
134 YsgInArgLysLysValIleMetGlnAlaPheSerArgLysAlaLeuGln 150
|||||
427 AGCAGCGAAGAAAGATGATTATTCAGGCTTCAGCGCGGACGACATCCAG 476
|||||
151 CysTyrValLeuValIleAlaGlnGluValSerSerCysLeuGlnTrp 167
|||||
477 TGTACAGCTGCTCGATCGCTGAGAAATCAGCAGATTGTCTGAGCAGTG 526
|||||
167 PheSerCysGlyLysArgGlyLeuLeuValTyrProGluValLysArgL 184
|||||
527 GCTAAGCTGGGCGAGCGGCTCTCGGTCTACCCCGAGGTAAACGGCC 576
|||||
184 euMetPheArgIleAlaMetArgIleLeuLeuGlyCysGluProGlyPro 200
|||||
```

```
577 TCATCTCCGCAATCGCCATGGCCATCGCATCTGCGGCTGCGAGCGGGTCCA 626
201 AlaGlyGlyGlyLysGluAspGluGlnLeuValGluAlaPheGluGluLe 217
|||||
627 GCGGGCGGGGGGAGGAGCAGCAACAGCTCGTGAGGCTTTGAGAGAGAT 676
|||||
217 tThrArgAsnLeuPheSerLeuProIleAspValIleProPheSerGlyLeu 234
|||||
677 GACCGCATCTCTTCTCTCTTCCATTGACGTGCGCCCTTAGGGGCTCTGT 726
|||||
234 YrArgGlyValLysAlaArgAsnLeuIleHisAlaArgGluGluAsn 250
|||||
727 ACCGGGCGTGAAGCGCGGAACTTATACACCGCGCATCGAGGAGAAC 776
|||||
251 IleArgAlaLysIleArgArgLeuGlnAlaThrGluProAspGlyGlyC 267
|||||
777 ATTCGCGCCAGATCCGCGGCTTCAGGCTACAGACCGGATGGGGTTG 826
|||||
267 sLysAspAlaLeuGlnLeuLeuIleGlnHisSerTrpGluArgGlyGlu 284
|||||
827 CAAGACGCGCTGAGGCTCTGATTGAGCACTCGTGAGAGAGGAGAGAGA 876
|||||
284 rGluLeuAspMetGlnAlaLeuLysGlnSerSerThrGluLeuLeuPheGly 300
|||||
877 GCGTGATATGCAAGGCACTAAACAAATCTCAACAGAGCTCTCTTGGT 926
|||||
301 GlyHisGluThrThrAlaSerAlaAlaThrSerLeuIleThrTyrLeuG 317
|||||
927 GGTATGAACCTACAGCAGAGCTGCGACATCACTGATCTTACCTAGG 976
|||||
317 YLeuTyrProHisValLeuGlnLysValArgGluGluIleLysSerLysG 334
|||||
977 ACTTACCCACATGCTCTCCAGAAAGTTCCGAAAGATGAAGAGCAAGC 1026
|||||
334 YLeuLeuCysLysSerAsnGlnAspAsnLysLeuAspMetGluThrLeu 350
|||||
1027 GCTTACTTTCGAAGACAAATCAAGCAACAAAGTAAACATGGAACCTTGG 1076
|||||
351 GluGlnLeuLysTyrIleGlyCysValIleLysGluThrLeuArgLeuAs 367
|||||
1077 GACACGCTTAATATCATTTGGGTGTGTATTAAGAGAACCCGCGATTGAA 1126
|||||
367 nProProValProGlyGlyPheArgValAlaLeuLysThrPheGluLeu 384
|||||
1127 TCTCCGGTCCAGAGAGGTTTCGGGTTGCTCTGAGACTTTTGAAGCTGA 1176
|||||
384 snGlyTyrGlnIleProLysGlyTrpAsnValIleTyrSerIleCysAsp 400
|||||
1177 ATGATATCCAAATCCCAAGGCGTGGAAATGTTATTTACAGTATCTGTGAC 1226
|||||
401 ThrHisAspValAlaAspIlePheThrAsnLysGluGluPheAsnProAs 417
|||||
1227 ACCCAGATGTGGCAGATATCTTCACTAACAAGAGAAATTAATCCGGA 1276
|||||
417 PArgPheIleValProHisProGluAspAlaSerArgPheSerPheIle 434
|||||
1277 CCGCTTATAGTGCTCATCCAGAGATGCTCCCGGTTCACTTCATATTC 1326
|||||
434 roPheGlyGlyLysLeuArgSerCysValGlyLysGluPheAlaLysL 450
|||||
1327 CATTTGAGAGAGCCCTTCGAGCTGTGTAGCAAGAAAGATTGGCAAAATTT 1376
|||||
451 LeuLeuLysIlePheThrValGluLeuAlaArgHisCysAspTrpGln 467
|||||
1377 CTTCTTAAGATTTTACAGTGGAGCTGGCTGAGCACTGTGATTTGGCAGCT 1426
|||||
467 uLeuAsnGlyProProThrMetLysThrSerProThrValTyrProVal 484
|||||
1427 TCTAATATGACCTCTACAAATGAAGAAAGCCCACTGTATACCTGTGG 1476
|||||
484 sPAsnLeuProAlaArgPheThrTyrPheGlnGlyAspIle 497
|||||
1477 ACAATCTCCCTGCAAGATTCACTTCTCAGGAGATATTC 1517
|||||
```

seq_name: gb_ptr4:AF005418
seq_documentation_block:
LOCUS AF005418 1743 bp mRNA PRI 15-DEC-1997
DEFINITION Homo sapiens retinoic acid hydroxylase mRNA, complete cds.
ACCESSION AF005418
VERSION AF005418.1 GI:2688845
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS White,J.A., Beckett-Jones,B., Guo,Y.D., Dilworth,F.J., Bonasoro,J.,
Jones,G. and Petkovich,M.
TITLE cDNA cloning of human retinoic acid-metabolizing enzyme (hP450RA1)
IDENTIFIES a novel family of cytochromes P450
J. Biol. Chem. 272 (30), 18538-18541 (1997)
MEDLINE 97373542
REFERENCE 2 (bases 1 to 1743)
AUTHORS White,J.A., Beckett-Jones,B., Guo,Y., Dilworth,F.J., Bonasoro,J.,
Jones,G. and Petkovich,M.
TITLE Direct Submission
JOURNAL Submitted (26-MAY-1997) Cancer Research Labs, Queen's University,
Botterell Hall, Rm 355, Kingston, Ont K7L 3N6, Canada
FEATURES
source location/Qualifiers
1..1743
/organism="Homo sapiens"
/db_xref="taxon:9606"
39..1532
/note="hP450RA1: cytochrome P450"
/product="retinoic acid hydroxylase"
/protein_id="AA888881.1"
/db_xref="GI:2688846"
/translation="MGPLPILASALCTFVLPPLFLFAIKIMDYCVSGRDRSCALPL
PFGTGVPEFEETLQWYIQRRKFLQMKRRKGYLTHLFGPRVYRWGADNVRILL
GDDRVSVWMPASVKTILSGCLSNLHDSKROKRYIMRARSFALECYAPITEEV
GSLDQMLSCGERGLYPERKRLMRFLAMILLGCEPGLAGDSBOIYVAFPEET
RNLFLPDIYFESGLYRGKARNLIFARIQNTIRATIGCLARSSEAGCCDADLQILE
HMERGERLMDQALIKOSTELLFGHERTSASLSITTLGLTPIVLOKREELKSG
LLCKSNQNKIMELIEQIKTYIGVYKTLNLPVPVGGFVALKTELNGYQIPKGM
NVTISCDTHVAIEFTNKEEFNPREMLPHPDASREFSIFPGGLRSCVGEFEAKI
LKRTFVLAHCDQMLNGPMTKTSPTYYPVNDLPAFTHFGEI"
BASE COUNT 423 a 442 c 477 g 401 t
ORIGIN
alignment_scores:
Quality: 2441.00 Length: 497
Ratio: 5.033 Gaps: 0
Percent Similarity: 97.586 Percent Identity: 93.360
alignment_block:
US-09-668-482-32 x AF005418 ..
Align seg 1/1 to: AF005418 from: 1 to: 1743
1 MetGlyLeuProAlaLeuLeuAlaSerAlaLeuCyThrPheValLeuPr 17
39 ATGGGGCTCCGGGGCTGGCGCATGCTGGACCTTGTCGCTGCC 88
17 oleuLeuPheLeuAlaLeuLysLeuTrpAspLeuTrpCysValS 34
89 GCTGCTCTTCTTCTGCTGCGATCAAGCTCTGGGACCTGACTGCTGA 138
34 eSerArgAspArgSerCysAlaLeuProLeuProProGlyThrMetGly 50
139 GCGGCGCGGACCGACATTGCGCTCCCATTTGCCCGGGGACTATGGGC 188
51 PheProPhePheGlyLthrLeuGlnMetValLeuGlnArgArgLysPh 67
|||||

189 TTCCTCTCTTGGGAAACCTTGCAGATGTACTGCAGCGGAGAACTT 238
67 eLeuGlnMetLysArgArgLysTrpGlyPheLleTrpLysThrHisLeuP 84
239 CTGCGAGATGAAGCGGAGAAATACGGCTTACTTACAAAGACGACTGT 288
84 heGlyArgProThrValArgValMetGlyAlaAspAsnValArgArgL 100
289 TCGGGCGGCGCCACCGTACGGGTATGGGGCGGAGCAATGTGGCGCATC 338
101 LeuLeuGlnGlnHisArgLeuValSerValHisTrpProAlaSerValAr 117
339 TTCTTGAGAGACCGGCTGTGTGGGTCCACGTGGCAGCCGTGTCGCG 388
117 gThrIleLeuGlnValaGlyCysLeuSerAsnLeuHisAspSerSerHisL 134
389 CACCATTCGGGATCTGCTGCTCTCTACCTGACGACTCTCTGACACA 438
134 ySerLArgLysLysValIleMetGlnAlaPheSerArgLValaLeuGln 150
439 AGCAGCGCAAGAGAGTATATGCGGCTTCACGCCGAGCACTCGAA 488
151 CysTrpValLeuValIleAlaGlnGlnValSerSerCysLeuGlnGlnTr 167
489 TGCTTAGCTGCCGGTATCACCGAAGTGGGACGACCTGGAGCACTG 538
167 PleuSerCysGlyLthrArgGlyLeuLeuValTrpProGlnValLysArgL 184
539 GGTGAGCTGCGCGGAGCGCGGCTCTGCTACCCGAGTGAAGCGCC 588
184 euePheArgIleAlaMetArgIleLeuLeuGlyCysGlnProGlyPro 200
589 TCATGTTCCGATGCCATGCGATCTTACGTGGCTGCGCAACCCACACTG 638
201 AlaGlyGlyGlyLysAspGlnGlnLeuValGlnAlaPheGlnGlnL 217
639 GCGGCGGAGCGGCGACTCCGAGCAGCACTGTGGAGGCTTGAGGAAT 688
217 ThrArgAsnLeuPheSerLeuProIleAspValProPheSerGlyLeuT 234
689 GACCGCAATCTCTCTGCTGCCATCGACGTGCCCTTCACCGGCGCTGT 738
234 yArgGlyValLysAlaArgAsnLeuIleHisAlaArgIleGlnGlnAsn 250
739 ACCGGGCGATGAAGCGCGGAACTCATTCACCGCGCATCGAGAGAAC 788
251 IleArgAlaLysIleArgArgLeuGlnAlaThrGluProAspGlyGly 267
789 ATTCCGCGCAAGATCTGCGGCTGCGGCGATCCGAGCGGCGAGGCTG 838
267 sLysAspAlaLeuGlnLeuLeuIleGlnHisSerTrpGluArgGlyL 284
839 CAAGAAGCGGCTGACGCTGTATGAGCACTGTGGGAGAGGAGAGC 888
284 rGLeuAspMetGlnAlaLeuLysGlnSerSerThrGluLeuLeuPheGly 300
889 GCGTGACATGACGACCTAAAGCAATCTTCACAGCAACTCTCTTTGGA 938
301 GlnHisGlnThrThrAlaSerAlaAlaThrSerLeuIleThrTrpLeuG 317
939 GGACACGAACACAGCGCGAGTGCAGCCACATCTGTGATCACTTACCTG 988
317 yLeuTrpProHisValLeuGlnLysValArgGlnGlnLleLysSerLys 334
989 GCTTACCCACATGTTCTCCAGAAAGTGCAGAAAGCTGAAGAGTAAAG 1038
334 yLeuLeuCysLysSerAsnGlnAspAsnLysLysLeuAspMetGluThrLeu 350
1039 GTTACTTTCAGAGACATCAAGCAACAAAGTTGAGACATGAATTTTG 1088
351 GlnGlnLeuLysTrpTrpIleGlyCysValIleLysGlnThrLeuArgL 367
1089 GAACACTTAATACATCGGGGTGTATTAAAGAGACCTTGCACGTGAA 1138

367 nProProValProGlyGlyPheArgValAlaLeuLysThrPheGluLeuA 384
 |||||
 1139 TCOCACATTCAGAGAGGTTGCGGTGCTCTGACAGACTTTGAATTA 1188
 384 sngLYTyrGlnIleProLysGlyTTPAsnValIleTyrSerIleCysAsp 400
 |||||
 1189 ATGGATACCAATTCACCAAGGCTGGAAATGTTATCTACAGTATCTGTGAT 1238
 401 ThrHisAspValAlaAspIlePheThrAsnLysGluGluPheAsnProAs 417
 |||||
 1239 ACTCATGATGTCGACAGATCTTCACCAACAAGAGAAATTAATTCGGA 1288
 417 ParPheIleValProHisProGluAspAlaSerArgPheSerPheIleP 434
 |||||
 1289 CCGATTCATGCTGCTCACCACAGAGATGATCCAGTTCAGCTTCATTC 1338
 434 rOPheGlyGlyLeuArgSerCysValGlyLysGluPheAlaLysIle 450
 |||||
 1339 CATTTGGAGGAGGCTTAGAGCTGTGAGCAAGAAATTTGCAAAATTT 1388
 451 LeuLeuLysIlePheThrValGluLeuAlaArgHisCysAspTTPGlnLe 467
 |||||
 1389 CTTCACAAATATTTACAGTGCAGCTGCAGGCATTTGCACTGCACACT 1438
 467 uLeuAsnGlyProProThrMetLysThrSerProThrValTyrProValA 484
 |||||
 1439 TCTAAATGACCTCCTACAAATGAAAAACAGTCCACCGTATCTGTGG 1488
 484 sPAsnLeuProAlaArgPheThrTyrPheGlnGlyAspIle 497
 |||||
 1489 ACAATCTCCTCGAAGATTCACCCATTTCCATGGGGAATTC 1529

seq_name: gb_ov:AF199462

seq_documentation_block:

LOCUS AF199462 1479 bp mRNA VRT 21-NOV-1999
 DEFINITION Gallus gallus retinoic acid degrading enzyme CYP26 mRNA, complete cds.

ACCESSION AF199462
 VERSION AF199462.1 GI:6456740

KEYWORDS
 SOURCE chicken.

ORGANISM Gallus gallus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;

REFERENCE
 1 (bases 1 to 1479)
 Swindell,E.C., Thaller,C., Sockanathan,S., Petkovich,M.,

AUTHORS Jessell,T.M. and Eichele,G.

TITLE Complementary Domains of Retinoic Acid Production and Degradation

JOURNAL In the Early Chick Embryo

REFERENCE Dev. Biol. (2000) In Press

AUTHORS Swindell,E.C., Thaller,C., Sockanathan,S., Petkovich,M.,

TITLE Direct Submission

JOURNAL Submitted (28-OCT-1999) Max Planck Institute for Experimental

FEATURES
 Endocrinology, Feodor-Lynen Strasse 7, Hannover 30625, Germany

location/Qualifiers
 1..1479

source
 /organism="Gallus gallus"

CDS
 /db_xref="taxon:9031"

1..1479
 /note="P450RA1"

/codon_start=1
 /product="retinoic acid degrading enzyme CYP26"

/protein_id="AAFO9250.1"

/db_xref="GI:6456741"

/translation="MGSAIVASALCTPLPLFLIAVRIMDLTCASGRDPSGLPL
 PEGTGLPFGSETLQVLRKFLQMKRRKIGFYIKHLRPPYRVYAGANVHILL
 GEHRLVSVQWASVFTILISGCLSLHNGHRRKRVYQAFSDALQHTVPIQEEY
 SACLQWLTGAGPCLLVPEVRLMEFRIAMRILTGFPQASPDDEQQLVEAFEMIRN

BASE COUNT 302 a 438 c 427 g 312 t
 ORIGIN
 LPSLDVPSFSGIRGLRARNIIIAKIEINIRAKREBEQKDALQILMERTQGN
 GEOLNMOIKRESATEILFGHEHTTASATSLIAFLGHVLOKVRKLOKGLLSP
 NOELNMEFLEQLKYTCVKETRLRSPFPGFRALAKTLFELNGYQIRKGMVVIS
 ICDHADVADLTDDDEFNDRFMSPEDESRSEFIPGGGLRSCVKEKRAKVLKIF
 TVEIARSCDMQLNGPPMTKGTPIYVPDNLPAKFIQSGOI"

alignment_scores:
 Quality: 2057.50 Length: 497
 Ratio: 4.502 Gaps: 3
 Percent Similarity: 91.952 Percent Identity: 80.080

alignment_block:
 US-09-668-482-32 x AF199462 ..

Align seg 1/1 to: AF199462 from: 1 to: 1479

1 MetGlyLeuProAlaLeuAlaSerAlaLeuCysThrPheValLeuPr 17
 |||||
 1 ATGGGCTTCTCCGCTGTGGTGGCCAGCGCCCTGTGCACCTTCCTGCTGCC 50
 17 oLeuLeuLeuPheLeuAlaAlaLeuLysLeuTyrPAspLeuTyrCysValS 34
 |||||
 51 CTTGCTGCTCTCTCTGGCGCCGCTCGCGCTCTGGAGCTGACTGCGCCA 100
 34 eSerArgAspArgSerCysAlaLeuProLeuProProGlyThrMetGly 50
 |||||
 101 GCGGCGCGGACCCGACGCTGCTCCGCGGTGGCCCGGCGACCATGGGG 150
 51 PheProPhePheGlyGluThrLeuGlnMetValLeuGlnArgArgLysph 67
 |||||
 151 CTGCGCTTCTTCCGGGAGACGCTGACAGATGCTGCGACGCGCGGAAT 200
 67 eLeuGlnMetLysArgArgLysTyrGlyPheIleTyrLysThrHisLeuP 84
 |||||
 201 CCGAGATGAGAGCGAGAGAAATAGCGCTCATCTACAGACTCACTCT 250
 84 heGlyArgProThrValArgValMetGlyAlaAspAsnValArgArgIle 100
 |||||
 251 TTGGCGCGCCGACGCTGCGGTGAGGGGCTGAGAAATGTGCGCACATC 300
 101 LeuLeuGlyGlnHisArgLeuValSerValHisTyrProAlaSerValAr 117
 |||||
 301 CTGCGGGCGAGACCGGCTGTCTCGTGCAGTGGCCCGCTCGGCGG 350
 117 gThrIleLeuGlyAlaGlyCysLeuSerAsnLeuHisAspSerHisL 134
 |||||
 351 CACCATCTTGGGCTCGGCTGCTCCACACCTTCACAAACGGGAGACACA 400
 134 ySGLArgLysLysValIleMetGlnAlaPheSerArgGluAlaLeuGln 150
 |||||
 401 AACACCGAANAAGGTGATCATCGAGGCTTCTCCGGAGATCCCTGCAG 450
 451 CACTACAGGCGCTGCTCATCCAGAGAGGTGAGTGCCTCGTGCAGACAGT 500
 151 CysTyrValLeuValIleAlaGluLysValSerCysLeuGluGlnTr 167
 |||||
 451 CACTACAGGCGCTGCTCATCCAGAGAGGTGAGTGCCTCGTGCAGACAGT 500
 167 pleuSerCysGlyLysArgGlyLeuLeuValTyrProGluValLysArgL 184
 |||||
 501 GCTGGGTGCGGGGCC...TGCCTGCTGTATCCAGAGGAAGCGC 547
 184 eumetPheArgIleAlaMetArgIleLeuLeuGlyCysGluProGlyPro 200
 |||||
 548 TATGTTCCGCAATCGCATCGCATCTGCTGCGTTCACAGCCCGCCAG 597
 201 AlaGlyGlyGlyLysPglGlnGlnLeuValGluAlaPheGluGlu 217
 |||||
 598 GCC...AGCCCGATGATGAGCAGCAGCTGTGTGAGGCTTCGAGAGAT 644
 217 thrArgAsnLeuPheSerLeuProIleAspValPProPheSerGlyLeuT 234
 |||||

```

645 GATCCGCAACCTCTTCTCCCTCCCATCGATGCTTCACTGAGGCTCT 694
234 YTAAGGlyValLysAlaArgAsnLeuIleHisAlaArgIleGluGluAsn 250
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
695 ACCGGGGCTGGGGGGCGCAACATCATCATCCCAAGATGAGGGAAC 744
251 ILeaArgAlaLysIleArgArgLeuGlnAlaThrGluProAspGlyGlyCy 267
745 ATCCGTGCCAAGATGGCCCGC.....AAGGACCCCGAGGCTGCTA 785
267 slyAspAlaLeuGlnLeuLeuIleGluHisSerTrpGluArgGlyGluA 284
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
786 CAAGATGACATGACAGCTGTGTGATGAGCACACAGAGGTAATGAGAGAC 835
284 rLeuAspMetGlnAlaLeuLysGlnSerSerThrGluLeuLeuPheGly 300
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
836 AGCTCAACATGACAGAGCTGAGAGAGTGCACACAGAGCTGCTCTTTGA 885
301 GlyHisGluThrThrAlaSerAlaAlaThrSerLeuIleThrTyrLeuG1 317
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
886 GGTCAAGAACACAGCGCCAGTGTCTACATCATCATGATTGCTTCTAGG 935
317 yLeuThrProHisValLeuGlnLysValArgGluLeuLysSerLysG 334
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
936 GCTCCACCATGATGTTCTGCAGAAAGTGAGAAAGAGCTGCATTAAGG 985
334 yLeuLeuCysLysSerAsnGlnAspAsnLysLeuAspMetGluThrLeu 350
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
986 GGTACTCTCTGGTCCCAACCAAGAAAACAATGTAACATGAGGTTCTTG 1035
351 GluGlnLeuLysTyrIleGlyCysValIleLysGluThrLeuArgLeuAs 367
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1036 GAGAGCTGAGAGTACACAGCGTGTCTATCAAGAAACCTCAGCGCTAG 1085
367 nProProValProGlyGlyPheArgValAlaLeuLysThrPheGluLeuA 384
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1086 CCGCTCTCTTCTCGAGGATTCGGAATGCACTCAAGACCTTGAGCTGA 1135
384 snGlyTyrGlnIleProLysGlyTrpAsnValIleTyrSerIleCysAsp 400
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1136 ATGCTTACCAAGATCCCAAGCGTGAATGTTATTTCAGTATCTGCAT 1185
401 ThrHisAspValAlaAspIlePheThrAsnLysGluGluPheAsnProAs 417
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1186 ACCCACATGCGCCGATCTCTTACCGACAAGACGAGTCAACCCGGA 1235
417 PArgPheIleValProHisProGluAspAlaSerArgPheSerPheIleP 434
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1236 CCGCTTATGTCCTCCGTCCTCGAGGACTCTCTAGGTTCAAGTTTCAT 1285
434 rPheGlyGlyGlyLeuArgSerCysValGlyLysGluPheAlaLysIle 450
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1286 CTTTCGGTGGGGGCTGAGGAGCTGCGTGAAGCAAAAGATTGCAAAAGTC 1335
451 LeuLeuLysIlePheThrValGluLeuAlaArgLysCysAspTrpGlnLe 467
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1336 CTTCTAAAAAATTTTACGCTGAGAGCTGCGGAGCTGCTGACTGCGAGCT 1385
467 uLeuAsnGlyProProThrMetLysTrpSerProThrValTyrProValA 484
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1386 GCTAAATGAGACCTCTACATGAAAGGGGCCCATATGTAACCTCTGG 1435
484 sPasnLeuProAlaArgPheThrTyrPheGlnGlyAspIle 497
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1436 ACAATGCTGCTGCCAAATTCATAGTTTACGCGGCAAAATC 1476
seq_name: gp_hvg21:AL358613

```

```

VERSION AL358613.8 GI:13660841
KEYWORDS HTG; HTGS_Phrase1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE 1 (bases 1 to 166337)
JOURNAL Plumb,B.
COMMENT Direct Submission
Submitted (13-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone
requests: clonerequests@sanger.ac.uk
On Apr 17, 2001 this sequence version replaced gi:13625010.
Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: DA348J12
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; 108752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 161818 bases at least Q40
Consensus quality: 163957 bases at least Q30
Insert size: 165037; sum-of-contrigs
Insert size: 176845; 5.5% error; agarose-fp
Quality coverage: 5.93x in Q20 bases; sum-of-contrigs quality
coverage: 5.81x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 14 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 32409: contig of 32409 bp in length
32410 32509: gap of 100 bp
32510 52913: contig of 20404 bp in length
52914 53013: gap of 100 bp
53014 55817: contig of 2804 bp in length
55818 55917: gap of 100 bp
55918 74512: contig of 18595 bp in length
74513 74612: gap of 100 bp
74613 83833: contig of 9221 bp in length
83834 83933: gap of 100 bp
83934 87001: contig of 3068 bp in length
87002 87101: gap of 100 bp
87102 97488: contig of 10387 bp in length
97489 97588: gap of 100 bp
97589 100583: contig of 2995 bp in length
100584 100683: gap of 100 bp
100684 105295: contig of 4612 bp in length
105296 105395: gap of 100 bp
105396 129702: contig of 24307 bp in length
129703 129802: gap of 100 bp
129803 132462: contig of 2660 bp in length
132463 132562: gap of 100 bp
132563 137099: contig of 4537 bp in length
137100 137199: gap of 100 bp
137200 158309: contig of 21110 bp in length
158310 158409: gap of 100 bp
158410 166337: contig of 7928 bp in length.
Location/Qualifiers
1..166337
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="Rp11-348J12"

```

FEATURES

source

```

seq_documentation_block:
LOCUS AL358613 166337 bp DNA HTG 14-APR-2001
DEFINITION Homo sapiens chromosome 10 clone Rp11-348J12, ** SEQUENCING IN
PROGRESS ***, 14 unordered pieces.
ACCESSION AL358613

```

```
misc_feature      /clone.lib="RCF1-11.2"
1..32409
/note="assembly_fragment:02593
fragment_chain:1
clone_end:17
vector_side:left"
misc_feature      32510..52913
/note="assembly_fragment:01673
fragment_chain:1"
53014..55817
/note="assembly_fragment:01466
fragment_chain:2"
misc_feature      55918..74512
/note="assembly_fragment:01282
fragment_chain:2"
74613..83833
/note="assembly_fragment:02258
fragment_chain:2"
misc_feature      83934..87001
/note="assembly_fragment:00938
fragment_chain:2"
87102..97488
/note="assembly_fragment:01557
fragment_chain:2"
misc_feature      97589..100583
/note="assembly_fragment:00199
fragment_chain:2"
100684..105295
/note="assembly_fragment:01309"
105396..129702
/note="assembly_fragment:02028.0"
129803..132462
/note="assembly_fragment:02077"
132563..137099
/note="assembly_fragment:02005
fragment_chain:3"
137200..158309
/note="assembly_fragment:00282
fragment_chain:3"
158410..166337
/note="assembly_fragment:01837
fragment_chain:3
clone_end:SP6
vector_side:right"
BASE COUNT      46344 a 33577 c 34990 g 50115 t 1311 others
ORIGIN
alignment_scores:
Quality: 1939.50      Length: 1166
Ratio: 4.007          Gaps: 13
Percent Similarity: 41.509      Percent Identity: 39.623
alignment_block:
US-09-668-482-32 x AL358613 ..
Align seg 1/1 to: AL358613 from: 1 to: 166337
1 MetGlyLeuProAlaLeuAlaSerAlaLeuCysThrPheValLeuPr 17
|||||
54914 ATGGGGCTCCCGCGCTGCTGCGATCAAGCTCGGACCTGACTGCTGCC 54963
17 OleuLeuLeuPheLeuAlaAlaLeuLysLeuTyrPaspLeuTyrCysValS 34
|||||
54964 GCTGTGCTCTTCCTGCTGCGATCAAGCTCGGACCTGACTGCTGCC 54963
34 eTSerArgPargSerCysAlaLeuProLeuProGlyThrMetGly 50
|||||
55014 GGGCGCGCAGCGCACTTGCTGCCCTCCCATTCGCCCGCGGACATGAGG 55063
51 PheProPhePheGlyGluThrLeuGlnMetValLeu.Gln..... 63
|||||
55064 TTCCTCTTCTGGGGAAACCTTGCAGATGATGACAGGTAAAGGAGG 55113
```

```
63 ..... 63
55114 TGGGGCGGAGACAGCTGTTCCCGAGACCGCGCGCTGGCTTCT 55163
63 ..... 63
55164 GCTGAAGTGGGGTAGAGCCCCCGGAGGCATGCTATTGGGCTAGAG 55213
63 ..... 63
55214 CAGGCTGGCGGAGCGCGGCTCCCGCCCCCTCATGCCCCACTTC 55263
64 ..... ArgArgLysPheLeuGlnMetLysArg 72
55264 TCTCTCCGCTTCTCTCCACACAGGAGGAGTCTCTGCGATGAAGCC 55313
73 ArgLysTyrGlyPheLeuTyrLysThrHisLeuPheGlyArgProThrVa 89
|||||
55314 AGGAATACGGCTTCATCTACAAGACGCACTGTGGCGGCCACCGT 55363
89 IArgValMetGlyAlaAspAsnValArgArgIleLeuLeuGlnHisA 106
|||||
55364 ACGGTGATGGCGCGGACAAATGTGCGCGCATCTTGCGAGAGACACC 55413
106 rGleuValSerValHisThrProAlaSerValArgThrIleLeuGlyAla 122
|||||
55414 GCTGGTGTGGTGCACATGGCCAGCTCGGTGGCGCACCATTTCTGGGATCT 55463
123 GLeuSerLeuSerAsnLeuHisAspSerSerHisLysGlnArgLys.Lys. 138
|||||
55464 GGTGCTCTCTTAACCTGCACGACTCTCTGCACACAGCGCAAGAGT 55513
138 ..... 138
55514 GGGGCGAGAGCGGAGGCTGTGACAGGAGGAGGACCCATTATGAGCG 55563
138 ..... 138
55564 GAATTCGGCTGATGATGCTAGGCGGCGGCTAGCAGCTTGAGTGGGCT 55613
138 ..... 138
55614 AGGACCTCTGCCAGCTCCAGGTTAGCTATACCCAGCTCGAGAGTGCA 55663
138 ..... 138
55664 TGTGTCTGCGAGACTGGGGGTGTGTGAAGGGAGCGCGGTAGACGAGA 55713
139 ..... ValI1 140
55714 GGGCGGATGAGGCTTTTAACGCTGTCCCTCTCGGAGCTCAGAGTAT 55763
140 eMetGlnAlaPheSerArgLysAlaLeuGlnCys..... 151
|||||
55764 TATCGGGCTTCAGCGCGAGGACACTGAAATGCTACGTCCGCGTATCA 55813
151 ..... 151
55814 CCGANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 55863
151 ..... 151
55864 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 55913
152 .... TyrValLeuValI1ealaglgluValSerSerCysLeuGlnT 167
|||||
55914 NNNNTACGTGGCGGTGATCACCGAGGAAGTGGGCAAGCAGCTGAGAGT 55963
167 rPLeuSerCysGlyGluArgGlyLeuLeuValTyrProGluValLysArg 183
|||||
55964 GCCTGAGCTGGCGGAGCGGCTCTCTGCTTACCCCGAGGTGAACGCG 56013
```

```

184 LeuMetPheArgIleAlaMetArgIleLeuLeuGlyCysGluProGlyLyr 200
|||||
56014 CTCATGTTCCGAATGCGCATGGCATCTCTACTGGGCTGCGAAACCCCACT 56063
200 CalAglyIylglYluAspGluGlnGluValGluAapheglum 217
|||||
56064 GGGGGGCGACGGGACTCCGAGCAGCAGCTTGTGGAGGCTTCGAGAAA 56113
217 eThrArgAsnLeuPheSerLeuProIleaspValProPheSerLylLeu 233
|||||
56114 TGACCCGGAATCTCTTCTGCTGCCATCGAGGtGtCCCTTCAGCGGGGCT 56163
234 TyrArg..... 235
|||||
56164 TACCGGGTAAGGGCGCAACGGGCTGGGACTAGGGCGGGGAGCTGG 56213
236 .....GlyValLysAlaArg 240
56214 GCGTCTGCTACCGCGCGGCTCTGCGCTCAGGGCATGAGCGCGCG 56263
241 AsnLeuIleHisAlaArgIleGluGluAsnIleArgAlaLysIleArg 257
|||||
56264 AACCTCATTCACGGCGGCGATGAGCAGAACATTGCGCCAAAGATCTGCGG 56313
257 gLeuGlnAlaThrGluProAspGlyGlyCysLysAspAlaLeuGlnLeu 274
|||||
56314 GCTCGGGGCATCCGAGGGCGGCGCATGCGCAAAAGACGCGCTGACGCTG 56363
274 euIleGluHisSerTrpGluArgGlyGluArgLysAspMetGln..... 288
|||||
56364 TGATCGAGACTGTTGGAGAGGGGAGAGCGGCTGACATGACAGTGAGT 56413
288 ..... 288
56414 AGCAGCTTCAGACCAAGCAGCTGGGAGTTTGTCTCCCTGGCTTCCAAAGC 56463
288 ..... 288
56464 GCTGTTCTGGGGCCCCAAAGCGCGCGCTGGGGCCAGCTTCTGGAG 56513
288 ..... 288
56514 TGGGGCGCGGCTCAGACTACAGCTATGGAATCCCGAAGAGGCTGAGA 56563
288 ..... 288
56564 CACCCGGTCAGAGAGCTGGGAAAGGGCTGGCGGGAAGCTGGAGAGATC 56613
288 ..... 288
56614 CCTAGCCTTCCAGGTTTCAAAAGGAAGTTGGAATTTGCAAAATGTT 56663
288 ..... 288
56664 AATTAAGAACCTTGGCATTTTAATAAATAAGACTTAAGCTAAGAGTT 56713
288 ..... 288
56714 TCCGCTAGAGCGGGGTCGTACTCGCTTACTGCTCCAGCTGAACATAAG 56763
288 ..... 288
56764 GACGTGTCATTTTGTAAAGATATGCTTCTCTGACTTCTGTGACGA 56813
288 ..... 288
56814 AATCATTTAGCCCTTCTAGTCTTCCCTCCAGAACTCTCAGTTGGAATTCG 56863
288 ..... 288
56864 AGTAATCTTGTGTCAACCGCAGGACACTTGTGAGAAATGTGGTCTCA 56913
289 .....AlaLeuLysGlnSerSerThrGluLeuLeuPheGly 301
|||||
56914 CTCATTCTTAGGCACTAAAGCAATCTTCAACGAACTCTCTTGGAG 56963
301 yHisGluThrThrAlaSerAlaAlaThrSerLeuIleThrTyrLeuGlyL 318
|||||
56964 ACAGAAACACAGGGCCAGTGCAGCACATCTCTGATCACTTACCTGGGG 57013
318 euTyrProHisValLeuGlnLysValArgGluGlnIleLysSerLys... 333
|||||
57014 TCTACCAATGTTCTCCAGAAAGTCGAGAGAGACTGAAGATAGGTA 57063
333 ..... 333
57064 GGGAGACTGGGCTGGGGGTGCTTATTAGCTTAGAAATTCAGCTGCT 57113
333 ..... 333
57114 CCGTAGCAACTCCGAATTAAGTCAGTGTGCTGCCTTCATGAGATATTT 57163
333 ..... 333
57164 GAAAGTCAAGGCCAGGGCTGCGTGGGCCAGTGGCAGAAATTAGCTTTG 57213
333 ..... 333
57214 TGAATAAACAGATGAGATCTTGAAGCTGTGAGCCACTTGGGAGGTTT 57263
333 ..... 333
57264 GACCTTGATTCTCAACCCCTCCACCTTTTGGTGAACCTGGAGATTTC 57313
333 ..... 333
57314 TCAGATAGTTCCACTTCTTGAAATGGTGTGTCACAGGCAATACATAGT 57363
333 ..... 333
57364 GTGAATAGCTGATTAGTGTGGTGTGTGTGATGTGGGGTGGGGTT 57413
333 ..... 333
57414 TTTTGGGGAACTTGTGCTTCTTAAGTGTGAGCAGCATTCCTGGGA 57463
333 ..... 333
57464 TTGTAATAGTATGATGATTGGCAGCAATGCCATTAGCTGTGTT 57513
333 ..... 333
57514 TCCCTTCCAGTCTCTCCATCATGGGGCCCTTGGGTTTATGCTCAC 57563
333 ..... 333
57564 TTAAAGCCCTGTTAAGTCTGCTGGGCTGATTTTATGGAGCACAAATAA 57613
334 .....GlyLeuLeuCysLysSerAs 340
57614 CTGTTCACCTCTGTATGACTGTTTGTATAGGTTTACTTTGCAAGACAA 57663
340 nGluAspAsnLysLeuAspMetGluThrLeuGluGlnLeuLysTyrIleG 357
|||||
57664 TCAAGACAAACAGTTGAGCATGGAATTTTGAACAACATTAATATACATCG 57713
357 LysValIleLysGluThrLeuArgLeuAsnProProValProGlyGly 373
|||||
57714 GGTGTATTATTAAGAGACCTTTCGACTGAATCCCGAGTTCCAGAGGG 57763
374 PheArgValAlaLeuLysThrPheGluLeuAsn..... 384
|||||
57764 TTTCGGGTTGCTGAAAGACTTTTGAATTAATGTAAATGTAATTAATCTCT 57813
384 ..... 384

```

57814 TCCTTCCCTTTTGTGTGCTTAAAACTCTCTTCTTCCTATGCT 57863
 384 384
 57864 GTGGGTGCATTTCTATGCTTTGTATATTTCTGCCCTATAGAGAT 57913
 384 384
 57914 AAGTTTCAGCAACCTGATCCACCTCTCTTCTACCTCCCTGCTT 57963
 384 384
 57964 TTAGCTCAATTTCCCAAGATATCAGTACTGCTTTTGTGTGA 58013
 385 Gly 385
 58014 AAGTAAATTCAGTTTGTACGCCCTTGCGTTTCATATCTTTGCAG 58063
 386 ..TyrGlnIleProLysGlyTTPAsnValIleTyrSerIleCysAspThr 401
 58064 GATACCAATTCCTCCAAAGGCTGGAAATGTTATCTACAGTATCTGTACT 58113
 402 HisAspValAlaAspIlePheThrAsnLysGluGluPheAsnProAsp 418
 58114 CATGATGTGGCAGAGATCTTCAACCAAGAGAAATTTATCTCGACCG 58163
 418 gPheIleValProHisProGluAspAlaSerArgPheSerPheIleProp 435
 58164 ATTCAATGCTGCCTCACCAGAGATGATCCAGTTTCAGCTTCATTCAT 58213
 435 heGlyGlyGlyLeuArgSerCysValGlyLysGluPheAlaLysIleLeu 451
 58214 TTGGAGGAGGCTTGTAGAGCTGTGTAGCAAAAGATTGCAAAATTTCTT 58263
 452 LeuLysIlePheThrValGluLeuAlaArgHisCysAspTyrGlnLeu 468
 58264 CTCAAATATTTACAGTGTGAGCTGGCCAGGCAATTGTGACTGGCAGCTCT 58313
 468 uAsnGlyProProThrMetLysThrSerProThrValTyrProValAsp 485
 58314 AATGAGACTCTTACATGATAAACCAGCTCCACCGTATCTCTGTGACA 58363
 485 snLeuProAlaArgPheThrTyrPheGlnGlyAspIle 497
 58364 ATCTCCTGCAAGATTCACTTCATGGGGAATC 58401
 seq_name: gb_af057566
 seq_documentation_block: 1479 bp mRNA VRT 10-MAR-1999
 LOCUS AF057566 DEFINITION Xenopus laevis retinolic acid converting enzyme mRNA, complete cds.
 ACCESSION AF057566
 VERSION AF057566.1 GI:3283171
 KEYWORDS
 SOURCE African clawed frog.
 ORGANISM
 Xenopus laevis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
 Xenopodinae; Xenopus.
 1 (bases 1 to 1479)
 Hollemann, T., Chen, Y., Grunz, H., and Pieler, T.
 Regionalized metabolic activity establishes boundaries of retinolic
 acid signaling
 EMBO J. 17 (24), 7361-7372 (1998)
 JOURNAL MEDLINE
 REFERENCE 2 (bases 1 to 1479)
 AUTHORS Chen, J. L., Grunz, H., Panitz, F., Pieler, T., and Hollemann, T.
 TITLE Direct Submission
 JOURNAL Submitted (05-APR-1998) Developmental Biochemistry, University of
 Goettingen, Humboldtallee 23, Goettingen, Lower Saxony 37073,
 Germany
 FEATURES
 source 1..1479 Location/Qualifiers

CDS
 1..1479
 /note="RACE"
 /codon_start=1
 /product="retinolic acid converting enzyme"
 /protein_id="AAC25158.1"
 /db_xref="GI:3283172"
 /translation="MDVLTLTSLACLTALPILLLTAAKIMEVYCLRRDACCANPL
 PPGTGLPFGSETLQWVLRFRFLQVRSQYGRYKTHHRSPTVRVATGAENRQILM
 GEHKLVSWMPASVNTILGACLSLNDHNEKRYTKVIAQFSREALANTVPQMEBY
 RCSVNLMTQSGCVLYVPAIKRMFRILAMRLILCCDQRMDEQETLLAEFEBSN
 LESLPIDVPSGLYRGLARNLIAQILENTEKLEKDEPDECKDALQQLIDSRNG
 EPINQALKESEATELLFGHGTTASATSLFSLAHKDVLEKRELETOGLSTRP
 BEKRLSTEIVQQLKYTSCVTKETRLSPVAGGFRAVAKTFVINGYQIIPKGMVYS
 IADTHGEADLPDDTDKENDRPLTLPDSSRPFIPGGGVRCICGERAKILIKVF
 VVELCRNDWELNGSPAMTTSPIICPDNLPAKFRPSSSI"
 BASE COUNT 350 a 422 c 384 g 323 t
 ORIGIN
 alignment_scores:
 Quality: 1758.50 Length: 498
 Ratio: 4.080 Gaps: 4
 Percent Similarity: 86.546 Percent Identity: 68.072
 alignment_block:
 US-09-668-482-32 x AF057566 ..
 Align seg 1/1 to: AF057566 from: 1 to: 1479

1 MetGlyLeuProAlaLeuLeuAlaSerAlaLeuCysThrPheValLeu 17
 1 ATGATCTGTACTCTGCTCTACACGCGCTGTGACCTCGACACGCC 50
 17 OleuLeuLeuPheLeuAlaLeuLysLeuTyrPAspLeuTyrCysVal 34
 51 ATTGCTGCTCTCTCAACCGCTGCCAAACTCTGGGAAGTCACTAGCTGA 100
 34 eSerArgAspArgSerCysAlaLeuProLeuProProGlyThrMetGly 50
 101 GCGCAAGATGCCGCGCTGCAACCCACGCCCCGGGACCATGGGA 150
 51 PheProPhePheGlyGluThrLeuGlnMetValLeuGlnArgArgLys 67
 151 CTGCCCTCTTGGAGAGACTCTGCAATGTGCTACAGAGCGCAGATT 200
 67 eLeuGlnMetLysArgArgLysTyrGlyPheIleTyrLysThrIleLeu 84
 201 CCTCCAGTGAAGCAAGCAATACGGTCCGATCTACAAAGCAGCATCTGT 250
 84 heGlyArgProThrValArgValMetGlyAlaAspAsnValArgArgIle 100
 251 TCGCAGGCCCCACGCGTGGCTACGGCGCGGAGAAAGCTTCGACAGATC 300
 101 LeuLeuGlyGluHisArgLeuValSerValHisTyrProAlaSerVal 117
 301 CTGATGGGGAGACACACTGTGTGCGTGTGACCTGGCGGCTCTGTGG 350
 117 gThrIleLeuGlyAlaGlyCysLeuSerAsnLeuHisAspSerSerHis 134
 351 CACCATCTCAGGGCGGCGTGTCTCCACCTGCACACAGACAGAGCACA 400
 134 yGlnArgLysLysValIleMetGlnAlaPheSerArgGluAlaLeu 150
 401 ACTTACACCAAGAGGTGATCGGCAAGCTTCTCCGGGGAAGCCCTGCT 450
 151 CysTyrValLeuValIleAlaGluLeuValSerSerCysLeuGluGln 167
 451 AATTACGTCCCGCAGATGGAAGAGGTGAGATCTCGGAGAACCTGTG 500
 167 pLeuSerCysGlyLysArgGlyLeuLeuValTyrProGluValLysArg 184

```

501 GCTGCAGACGGCCCC...TGTGTGCTGGTATATCCCGCATCAACGAA 547
184 eumetphearqtlleametarqllleuleuglyysgluproglypro 200
548 TATATGTCGCATATGCGATGAGGCTCTGCTGGCTGATATCC...CAA 594
201 AlaglyglylygluaspgluglnglueuValaAlaphhegluglume 217
595 CCATATGACCGCGAACAGAGGAAACGCTGCTGCAAGCCTTCGAGAAAT 644
217 tthrtargnleupheserleuproileaspyalPropheserglyleut 234
645 GAGCGGAAATCTTTCTCTTCCATATGATGCGCTTTAGCGGCTCT 694
234 YfargglYVallysaAlaargAsnleulleisAlaargllleglugluasn 250
695 ACCGGGCTTCGGGGCCAGAAACCTATTCATGCCCAATTAAGGAAAC 744
251 llearqalalyllleargargleuglnalathrgluproaspglyglycy 267
745 ATCAAGGAAAGCTGCAAGA...GAACGAGATGACATG 782
267 slsaspalaleuglnleullelglnhissertrpgluargglyglua 284
783 CAGGATGCTTCAGCTGCTGATTACTATAGCCGAGGAATGGCGAGC 832
284 rgleuaspmetglnalaleuylsglnserterhgluleuenehegly 300
833 CCATACATCTTCAGGATTAAGAGAGTCCGCAACGAGCTCTGTTGGA 882
301 glyhsgluthrthAlaserAlaathrsrleullethtyrleugl 317
883 GGCACGAGACGAGCAAGTGTCCAGCTCCACCTCTCTTCTTCTG 932
317 yleuTyrrProhlsValleuglnlyValarggluileuyserry 334
933 GCTTCACAAAGACGTTTCGAAAAAGTCCGACAGACCTTGAAACACAG 982
334 lyLeuleu...CyslysserAsnglnaaspaslnlyleuaspmetgluthr 349
983 GTCCTGTGCTCCACTAACCCGAGAAAGAAAGAACTGACATCGACGTA 1032
350 leugluglnleuyslyrllleglycysVallelysgluthrleuargle 366
1033 CTTCACAGCTTAAGATTAAGTACTAGCTGTCTATTAAGGACCTTCGGCT 1082
366 uasnProProvalProoglylypheargValaAlaleuysThrheglul 383
1083 CAGCCACACAGTGGCGGTGATTCGCGTCTCAAGACGTTTGTAT 1132
383 euaenglytyrGlnlleProlysglytrpAsnValleTyrrSerlleCys 399
1133 TAAATGGCTATACAGATTCCTCAAGGCTGGAATGTATATACAGATTCCT 1182
400 AspThrhisaspValaAlaaspIlephetrasnlysgluglupheaspr 416
1183 GACACCCACGCTGAGCGCATCTTTCCGACACGAGCAATTCACACC 1232
416 oAsparqPheleValProhlsProgluasAlaSerArpheserPheI 433
1233 TGAACGCTTCTTAACGCGACATTCACAGAGATTCCTCAAGTTTGGCTTCA 1282
433 leProheglyglylyleuargsercysValglylysgluPhaAlaLys 449
1283 TCCCTTTCGAGGTGAGTGTGCTGATCCGCAAGAAATTTGCTTAA 1332
450 lleuLeuLysIlePheThrValaGluleuAlaarghiscysAspTPgl 466
1333 ATCTCTTCAAGGTCTCTGCTGGAGATGTGTGTAATGCGACCTGGA 1382
466 nLeuLeuAsnglyProProthrhmetlyshrsrProthrhValtyrPro 483
1383 GCTCTGTAACGAGATCCCGCATGACACCTCCCGCATATATCTGTCTG 1432

```

```

483 alAspasnleuproAlaArqPheThrTyrrPheGlncllyaspile 497
1433 TGGCAACCTACCCGCTAAATTCAAACCTTCTCAGCTTATT 1476
seq_name: gb_ov:DRU68234
seq_documentation_block: 1479 bp mRNA VRT 22-NOV-1996
LOCUS DRU68234
DEFINITION Danio rerio all-trans-retinoic acid 4-hydroxylase (p450RAI) mRNA,
complete cds.
ACCESSION U68234
VERSION U68234.1 GI:1680715
KEYWORDS zebrafish.
SOURCE
ORGANISM Danio rerio.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Rasbora; Danio.
REFERENCE
1 (bases 1 to 1479)
White,J.A., Guo,Y.D., Baetz,K., Beckett-Jones,B., Bonasoro,J.,
Hsu,K.E., Dilworth,F.J., Jones,G. and Petkovich,M.
Identification of the retinoic acid-inducible all-trans-retinoic
acid 4-hydroxylase
J. Biol. Chem. 271 (47), 29922-29927 (1996)
97094702
JOURNAL
MEDLINE
REFERENCE
2 (bases 1 to 1479)
White,J.A., Guo,Y., Baetz,K., Beckett-Jones,B., Bonasoro,J., Hsu,K.
E., Dilworth,F.J., Jones,G. and Petkovich,M.
Direct Submission
Submitted (28-AUG-1996) Cancer Research Labs, Queen's University,
Botterell Hall, Kingston, ON K7L 3N6, Canada
FEATURES
source
1..1479
/organism="Danio rerio"
/db_xref="taxon:7955"
1..1479
/gene="p450RAI"
1..1479
/gene="p450RAI"
/note="all-trans-retinoic acid 4-hydroxylase"
/codon_start=1
/protein_id="AAC60045.1"
/db_xref="GI:1680716"
/translation="MGLITLMVFLCTIVPLVFLAATKIMLMIRRDPCRSPL
PGTMDLPFGTGLIILORRKLIRMKROKYGCTVTHLEFGRPTVYMGADNVRLL
GEHKLYVQMPASVRYTLLSDTISNVHGVQHKKKAKIMARSRLDLEHTIPVIOEV
KSAIQEWLQDSCVLYPEKKILMERIAMRIILGFPEEQIKDQELVEAFEEIMNL
FSLPIDVPSSGLRGLRARNFHSKIENIRKKIQQDDNENQYKDALQILTENSR
SDEPSIQAMKEAATFLFGGHETASTATSLVMFLGNTVEYOKRVEYQKVEVEMG
YTPGKLSMELDLQKYTGCVIKETLRIMPVPGFRVALKPEPLNGYQIPIKGMNVIY
SLCDDTVADVPNKPEFQPERMSGLEDGSPFNYPFGGGRMKCVGAEFAVLIKI
FLVELDHQCNWILSNPPTMKGPTIYPDNLPTKTYSTYRN"
BASE COUNT 425 a 331 c 379 g 344 t
ORIGIN
alignment_scores:
Quality: 1734.00 Length: 496
Ratio: 4.070 Gaps: 4
Percent Similarity: 85.887 Percent Identity: 67.742
alignment_block:
US-09-668-482-32 x DRU68234 ..
Align seg 1/1 to: DRU68234 from: 1 to: 1479
1 MetGlyLeuProAlaLeuLeuAlaSerAlaLeuCysThrPheValLeuPr 17
||||| ..|||
1 ATGGGGCTGTACACCTTATGGTCACTTCTGACACCATCGTCTACC 50
17 OleuLeuLeuPheLeuAlaAlaLeuLysLeuTTPasPleuTyrrCysVal 34
||||| ..|||

```

51 CGTTTACTACCTTTCTCCGCCGGCGTGAAGTCTGGGAGATCTTAATGATCC 100
34 *erSeratgspatgSerCysAlaLeuProLeuProGlyThrMetGly* 50
||| |||||
101 GAGCAGTCACTCCGAACTGACGAAGTCTCTACCGCAGGTACCATGAGGC 150
||| |||||
51 *pheProPhephcGlyThrThrLeuGlnMetValLeuGlnIleArgTrpGlysh* 67
||||| |||||
151 TTCCCGTTATTTGGAAACGCTCCAGCTGATCTCCAGAGAAAGACTT 200
||| |||||
67 *eLeuGlnMetLysArgArgLysTrpGlyPheIleTrpLysThrHisLeuP* 84
||| |||||
201 TCTGGCCATGAAGAACGGCAGAAATACGGGTGATCTCAAGACCACTCT 250
||| |||||
84 *heGlyArgProThrValArgValMetGlyAlaAspValAlaGlyIle* 100
||| |||||
251 TCGGGAACCCGACTGTCAGGGGTGATGGAGCTGTAATTTGAGAGCGATT 300
||| |||||
101 *LeuLeuGlyGlnHisArgLeuValSerValHisTrpProAlaSerValArg* 117
||| |||||
301 CTCTGTGGGGGAACACAGCTGGTGTCTGTCTCAGTGGCCACATCAGTAG 350
||| |||||
117 *gThrIleLeuGlyAlaGlyCysLeuSerAsnLeuHisAspSerSerHisL* 134
||| |||||
351 AACCATCTGGGTCTGACACCTCTCCAAATGTCATGAGGACTTCAACACA 400
||| |||||
134 *ysGlnArgLysLysValIleMetGlnAlaPheSerArgGlnAlaLeuGln* 150
||| |||||
401 AAAACAGAAAAAGGCATTAAGAGGGCTTCTCTCAGATGCTCTGGAG 450
||| |||||
151 *CysTrpValLeuValIleAlaGlnIleValSerSerCysLeuGlnGlnTr* 167
||| |||||
451 CACATCACTTCCCTGATTCACGAGAGGTGAAGAGCCCATACAGGAAATG 500
||| |||||
167 *PLEU.....SerCysGlyGlnArgGlyLeuLeuValTrpProGluV* 181
||||| |||||
501 GCTGCAAAAAAGACTCTGTC.....GTGCTGGTTATTCACAGAA 538
||| |||||
181 *AllyArgLeuMetPheArgIleAlaMetArgIleLeuLeuGlyCysGlu* 197
||| |||||
539 TGAAGAAATCATGTTTCGATAGCTAAGAAATCTGCTGGTTGTGA 588
||| |||||
198 *ProGlyProAlaGlyLysLysLysProGlnGlnGlnLeuValGlnIleAla* 214
||| |||||
589 CCAAGCAAAATA.....AAGACGAGCAGACGAAGAAGCTGGGAAAGCTTT 632
||| |||||
214 *eGlnGlnMetThrArgAsnLeuPheSerLeuProIleAspValProPheS* 231
||| |||||
633 TGAAGAAATGATCAAAAACCTGTCTCTGCGCAATGACAGCTCTTCA 682
||| |||||
231 *erGlyLeuTrpArgGlyValLysAlaIleArgAsnLeuIleHisAlaArgIle* 247
||| |||||
683 GTGCTCTGTACAGGGGTTTGAAGGACAGCAATTTCACTCCAAATTT 732
||| |||||
248 *GlnIleAsnIleLeuArgAlaLysIleLeuArgLeuGlnAlaIleArgIleProAs* 264
||| |||||
733 GAGGAAAAACATCGAGAAAGAAATTTCAAGT..GACGCACATGAAAACGA 779
||| |||||
264 *pGlyGlyCysLysAspAlaLeuGlnLeuLeuIleGlnHisSerTrpIleuA* 281
||| |||||
780 ACAGAAATACAAAGACGCCCTTACGCTGTTCATCGAAGAACAGCAGAGAA 829
||| |||||
281 *rgGlyGlnArgLeuAspMetGlnAlaLeuLysGlnSerSerThrGlnLeu* 297
||| |||||
830 GTGACGAACCTTTTAACTTTCACGGCATAAAGAGACAGCTTACAGAGCTT 879
||| |||||
298 *LeuPheGlyGlnHisGlnIleThrTrpAlaSerAlaIleThrSerLeuIleTh* 314
||| |||||
880 CTAATTTGAGGTGATGAACACACCGCACACTGCAACCTCACTTGTTCAT 928
||| |||||
314 *tTyIleuGlyLeuTrpProHisValLeuGlnLysValArgGlnGluIleL* 331
||| |||||
930 GTTCTTGCGTCTGAAACAGAACTGGTGTGAGAAAGTCTCAGAGAGAGGCTC 979

seq_name: gb_pf5:AF252297

```
seq_documentation_block:
LOCUS       AF252297             4445 bp             mRNA                    PRI             14-JUN-2000
DEFINITION  Homo sapiens cytochrome P450 retinoid metabolizing protein
            P450RA1-2 mRNA, complete cds.
ACCESSION   AF252297
VERSION     AF252297.1
KEYWORDS    GI:8515440
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE   1 (bases 1 to 4445)
AUTHORS    White,J.A., Ramshaw,H., Talmi,M., Stangle,W., Zhang,A.,
            Everingham,S., Creighton,S., Tam,S.-P., Jones,G. and Petkovich,M.
            Identification of the human cytochrome P450, P450RA1-2, which is
            predominantly expressed in the adult cerebellum and is responsible
            for all-trans-retinoic acid metabolism
            Proc. Natl. Acad. Sci. U.S.A. 97 (12), 6403-6408 (2000)
JOURNAL     Proc. Natl. Acad. Sci. U.S.A. 97 (12), 6403-6408 (2000)
MEDLINE     20300913
REFERENCE   2 (bases 1 to 4445)
AUTHORS    White,J.A., Ramshaw,H., Talmi,M., Stangle,W., Zhang,A.,
            Everingham,S., Creighton,S., Tam,S.-P., Jones,G. and Petkovich,M.
            Direct Submission
            Submitted (04-APR-2000) Cytochrome Inc., 116 Barrie Street,
            Bioscience Complex, Suite 2424, Kingston, ONT K7L 3N6, Canada
FEATURES
            location/Qualifiers
            1..4445
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            1..1539
```

CDS

```

/codon_start=1
/product="cytochrome P450 retinoid metabolizing protein
P450RAI-2"
/protein_id="AAF76003.1"
/db_xref="GI:8515441"
/translation="MLFEGIDVSAIALATLAAVLSTVLLAAVSOQLMRLMAATRDKS
CLPLIPKSGMGPFLIGETGMILLOGSFQSSREKREKPHILGRPLIRVAGANV
RKLIMGEHLVSTEMPRSTRMLGPTVSNSTGDIHRNKRKYSKIFSEHAEPSYLPK
IOLVITDILRAVSSHPEATINNYOEAKILFRNAIRVLLFSPFEDLGHLEFYQDFV
DNVFSIPVLPSPSGYRRIQAKIOLKGLKAIKREKLGCTQGDYLDALDLLESSKE
HKEMTMOLKDGTELELTPAAYATVASTSLIMOLKHPYLEKLRDELRAHGILHS
GGCPEGTLRLDLSGLRLDVIKEVMRLFTPISSGYRVLQTELDGFOIPIKGSV
MSIRDTHTAFVDVNFDFRSQSEDKDGFHYLPRFGGVRCTCGKHLAKLF
LKVLAVELASTSRFELATRTFRITLVPLHPVDGLSVKFFGLDSNQNLIPETEML
SAFTV"

```

```

BASE COUNT      893 a      1265 c      1240 g      1047 t
ORIGIN

```

alignment_scores:

```

Quality: 1048.50      Length: 495
Ratio: 2.841          Gaps: 9
Percent Similarity: 74.545      Percent Identity: 44.444

```

alignment_block:

```

US-09-668-482-32 x AAF252297 ..

```

```

Align seg 1/1 to: AAF252297 from: 1 to: 4445

```

```

7 LeuAlaSerAlaLeuCysThrPheValLeuProLeu.....Le 19
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
22 CTGGTGTGGCGGCGTGGCCACCGCTGCGCTGCTGCTGCTGCTGCGGCT 71
19 uLeuPheLeuAlaAlaLeuLysLeuTrpAspLeuTyrCysValSer 36
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
72 GCTGCTGCGCGCTGCTGCGACGACCTGCTGCGCTGCGCTGCGCGCAC 121
36 rGAspArgSerCysAlaLeuProLeuProGlyThrMetGlyPhePro 52
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
122 GCGCAAGAGACTGCAAGCTGCCATCCCAAGGATCCATGGGCTTCCG 171
53 PhePheGlyGluThrLeuGlnMetValLeuGlnArgArgLysPheLeu 69
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
172 CTGATGGAGAGACCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 221
69 nMetLysArgArgLysTyrGlyPheLeuTyrLysThrHisLeuPheGly 86
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
222 GTGCGGAGGAGAGATGAGTGCACACGTTCACAGACCATTTGTTGGGC 271
86 rGProThrValArgValMetGlyAlaAspAsnValArgArgLysLeu 102
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
272 GCGCGCTGATACCGCTGACCGGCGGAGAACGTGCGCAAGATCCATG 321
103 GlyGlnHisArgLeuValSerValHisTrpProAlaSerValArgThr 119
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
322 GGGGACACACACCTCTGTGACACCGAGTGGCTCGCAGCACCGCATGT 371
119 eLeuGlyAlaGlyCysLeuSerAsnLeuHisAspSerSerHisLysGln 136
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
372 GCTGGGCGCCCAACACGCTGTCCAAATTCATGGCAGATCCACCGACA 421
422 AGCGCAAGGCTTCTTCACAGATCTTCAGCCACGAGGCCCTGGAGATT 471
153 ValLeuValIleAlaGluGluValSerSerCysLeuGluGlnTrpLeu 169
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
472 CTGCCCCAAGATCCAGCTGGTGAATCCAGACACACTGCGCTGGAGCAG 521
169 rCysGlyGluArgGlyLeuLeuValTyrProGluValLysArgLeuMet 186
-|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
522 CCACCCCGAG...GCCATCAGCTGTACAGAGAGGCGCAAGAGCTACCT 568

```

```

186 heArgIleAlaMetArgLleLeuLeuGlyCysGlu...ProGlyProAla 201
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
569 TCCGATGGCCATCCGGGTGCTGCTGGGCTTCAGCATCCCT..... 609
202 GlyGlyGlyGluAspGluGlnGlnLeuValGluAlaPheGluGluMet 218
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
610 .....GAGGAGACACTGGGCACCTCTTACAGGCTACACGACATTGT 653
218 rArgAsnLeuPheSerLeuProIleAspValProPheSerGlyLeuTyr 235
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
654 GGAACAATGCTTCTCCCTGCTGCTGACCTCCCTTACGTGGCTACCGGC 703
235 rGlyValLysAlaArgAsnLeuIleHisAlaArgLleGluGlnAsnIle 251
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
704 GGGGATTCAGGCTGGCAGATCCTGCGAAGAGGCGTGAAGAAGCCCTC 753
252 ArgAlaLysIleArgArgLeuGlnAlaThrGluProAspGlyGlyCys 268
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
754 CGGGAGAAAGCTGCAGTGCACACAGGCAAGAGC.....TACTT 791
268 sAspAlaLeuGlnLeuLeuIleGlnHisSerTrpGluArgGlyGluArg 285
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
792 GGAGCCCTGGACCTCTCATTTAGAGCAGCAGAGAGCAGGAGGAGAGA 841
285 eAspMetGlnAlaLeuLysGlnSerSerThrGluLeuLeuPheGlyGly 301
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
842 TGACCATGCAAGACTGTAGAGCGGAGCCTGAGCTGATCTTGGCGCC 891
302 HisGluThrThrAlaSerAlaAlaThrSerLeuIleThrTyrLysGly 318
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
892 TATGCCACACGCGCAGCGCAGACCTCATCATGACCTGCTGAG 941
318 uTyrProHisValLeuGlnLysValArgGluGluIleLysSerLysGly 335
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
942 GCACCCACTGTGCTGGAGAGAGCTGCGGAGTACCTCGGCTCATGCA 991
335 euLeu.....CysLysSerAsnGlnAspAsnLysLeuAspMet 347
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
992 TCCGCAACATGGCGGCTGCCCTGC.....GAGGCAACACTGCGCTG 1035
348 GluThrLeuGlnGlnLeuLysTyrLleGlyCysValIleLysGluThr 364
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1036 GACACGCTGATGGGCTGCGCTGACTGACTGACTGACTGACTGACTG 1085
364 uArgLeuAsnProProValProGlyGlyPheArgValAlaLeuLysThr 381
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1086 GCGCTTTCACGCCCATTTCCGGCGGCTACCGCATCTGCTGACAGCT 1135
381 heGluLeuAsnGlyTyrGlnIleProLysGlyTrpAsnValIleTyrSer 397
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1136 TCGAGCTTGATGTTCCAGATGCCCAAGAGCTGAGTGCATATAGC 1185
398 IleCysAspThrHisAspValAlaAspIlePheThrAsnLysGluIup 414
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1186 ATCCGGACACCCATGACACAGCGCCGCTTCMAAAGCTGAACGTGT 1235
414 eAsnProAspArgPheIleValProHisProGluAspAla...SerArg 430
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1236 GAGCCCGATCGCTTACGCCAGCGCGGAGCAGAGCAAGAGATGGCGCT 1285
430 heSerPheIleProPheGlyGlyLeuArgSerCysValGlyLysGlu 446
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1286 TCCATTATCATCCCGTTGCGGCGGTGTCGAGCTGCTGGGCAAGCAC 1335
447 PheAlaLysIleLeuLeuLysIlePheThrValGluLeuAlaArgHis 463
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1336 CTGCGCAAGCTGTCTGAAGGTGCGGCTGCGTGCCTGCTGACACAG 1385
463 sAspTrpGlnLeuLeuAsnGlyPro...ProThrMetLysThrSerPro 479
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1386 CCGCTTGAAGCTGCTACACGAGACCTTCCCGCATCACCTTGTGCCCC 1435
479 hrValTyrProValAspAsnLeuProAlaArgPhe 490

```



```

346 pmetgluthrleugluGluLeuLysTyrIleGlyCysValIleLysGluT 363
      |||||||.....:||||:||||:||||:
63041 GCGAGAACTCTTAAATAAATGCCCTATCTAGACCAAGTGTACAGAAAG 63090
363 hrleuArgLeuasnProProValProGlyGlyPheArgValAlaLeuLys 379
      ::|||::||| ||||| ||||| ||||| |||||
63091 TATGCGGTATATCTCTCTGTGGCGGTGTTTCGAGAAATATATCA 63140
380 ThrPheGluLeuasnGlyTyrGluIleLeuProLysGlyTyrPasnValIleTyr 396
      ::|||::||| ||||| ||||| ||||| |||||
63141 GATGTCAATTTCAAGGCTTTTCATTTTCCCAAGGCTGTAGTTCTTAA 63190
396 rSerIleCysaspThrHisAspValAlaAspIlePheThrAsnLysGluG 413
      |||::||| ||||| ||||| ||||| |||||
63191 TCAAATTAGCCAAACACCGCGATCTGATTTATATCTGACCCAGAA 63240
413 lUpheAsnProAspArgPheIleValProHisProGluAspAlaSerArg 429
      |||::||| ||||| ||||| ||||| |||||
63241 AGTTTGATCCAGAACGGTTTACT.....CTGATGCTAGTGCACACT 63281
430 .....PheSerPheIleProPheGlyGlyGlyLeuArgSerCys 442
      |||::||| ||||| ||||| ||||| |||||
63282 CATATCTCCTCCCTTGCCCATGCTCTTGCGGCTTACGGGAATG 63331
442 sValGlyLysGluPheAlaLysIleLeuLeuLysIlePheThrValGluL 459
      |||::||| ||||| ||||| ||||| |||||
63332 TTTAGTAGAGGAATTTGCTCGGCTGGAATATATTTGCCACCCGAT 63381
459 euAlaArgHisCysAspTyrPglLeuLeuAsnGlyProPro...ThrMet 474
      |||::||| ||||| ||||| ||||| |||||
63382 TATTCACACATTTGATGACATTTGACCGGCGCAAAATTTAGAGTTG 63431
475 LysThrSerProThrValTyrProValAspAsnLeu 486
      |||::||| ||||| ||||| ||||| |||||
63432 GTTGTACCCCTAGTCCCGGCCCAAGATATTTG 63467
seq_name: gb_ov:AF185266
seq_documentation_block:
LOCUS AF185266 322 bp mRNA VRT 27-SEP-1999
DEFINITION Gallus gallus CYP26 (CYP26) mRNA, partial cds.
ACCESSION AF185266
VERSION AF185266.1 GI:5924309
KEYWORDS
SOURCE
ORGANISM
. chicken.
Gallus gallus
Archaeopteryx; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
1 (bases 1 to 322)
REFERENCE
1 Martinez-Ceballos,E. and Burdasa,J.C.A.
Direct Submission
Submitted (09-SEP-1999) Cell and Molecular Biology, Tulane
University, 2000 Stern Hall 6400 Freret St., New Orleans, LA 70118,
USA
FEATURES
Source
1..322 location/Qualifiers
/organism="Gallus gallus"
/db_xref="taxon:9031"
<1..>322
/gene="CYP26"
<1..>322
/gene="CYP26"
/note="contains heme binding domain; RA-metabolizing
enzyme"
/codon_start=2
/product="CYP26"
/protein_id="AAD56546.1"
/db_xref="GI:5924310"
/translatability="GERVALKTELENGYQIPKNNVYISTCDTHVDVADLFPTDKDEFNP
DREMSPPEDSSRPSFIPFGGLRSVCGKEPAKVLKIFTELARSCDWLLNGPPTM
KTSPT"

```

```

BASE COUNT 74 a 82 c 85 g 81 t
ORIGIN
alignment_scores:
Quality: 538.00 Length: 107
Ratio: 5.173 Gaps: 0
Percent Similarity: 97.196 Percent Identity: 90.654
alignment_block:
US-09-668-482-32 x AF185266 ..
Align seg 1/1 to: AF185266 from: 1 to: 322
373 GlyPheArgValAlaLeuLysThrPheGluLeuAsnGlyTyrGluIlePr 389
|||||
2 GGGTTTCGGGGTGTCTCTCAAGACCTTGAGCTGAATAGTTACCAATCCC 51
389 olYsgIYTPAsnValIleTyrSerIleCysAspThrHisAspValAla 406
|||||
52 CAAGGCTGGAGATGTTATTACAGTATCGCATACCAAGATGTGGCG 101
406 sPlePheThrAsnLysGluGluPheAsnProAspArgPheIleValPro 422
|||||
102 ATCTCTTACCGACAAGAGAGATTCAACCGGACGCTTCATATGTCCTCG 151
423 HisProGluAspAlaSerArgPheSerPheIleProPheGlyGlyLys 439
|||||
152 TCTCCGAGAGACTCTTAGTTCAATTCCTTGGTGGGGGCTT 201
439 uArgSerCysValGlyLysGluPheAlaLysIleLeuLysIlePheT 456
|||||
202 GAGAGACTCGTAGGCAAGAGTTTGCATAAAGCTCTTAAAAATTTTA 251
456 hrValGluLeuAlaArgHisCysAspTyrPglLeuLeuAsnGlyProPro 472
|||||
252 CGGTGAGACTGGCGGAGCTGTGACTGGCAGCTCTTAATATGACCTCG 301
473 ThrMetLysThrSerProThr 479
|||||
302 ACATGAAAAACAGCCCACT 322
seq_name: gb_p12:AF318211
seq_documentation_block:
LOCUS AF318211 1494 bp mRNA PLN 15-FEB-2001
DEFINITION Taxus cuspidata 5-alpha-taxadienol-10-beta-hydroxylase mRNA,
complete cds.
ACCESSION AF318211
VERSION AF318211.1 GI:12656591
KEYWORDS
SOURCE
ORGANISM
Taxus cuspidata.
Taxus cuspidata
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Coniferales; Coniferales; Taxaceae; Taxus.
1 (bases 1 to 1494)
REFERENCE
1 Schoendorf,A., Rither,C.D., Williams,R.M. and Croteau,R.B.
Molecular cloning of a cytochrome P450 taxane 10beta-hydroxylase
cDNA from Taxus and functional expression in yeast
Proc. Natl. Acad. Sci. U.S.A. 98 (4), 1501-1506 (2001)
2 (bases 1 to 1494)
Schoendorf,A. and Croteau,R.
Direct Submission
Submitted (30-OCT-2000) Institute of Biological Chemistry,
Washington State University, Pullman, WA 99164, USA
FEATURES
Source
1..1494 location/Qualifiers
/organism="Taxus cuspidata"
/db_xref="taxon:99806"
1..1494
/note="cytochrome P450-like protein"
/codon_start=1

```

```

/product="5-alpha-taxadienol-10-beta-hydroxylase"
/protein_id="AAK0946.1"
/db_xref="GI:12656592"
/translation="MDSFIRLSIGTFQGLSESPALISLTIPIAIIILLFRVH
KSSVLPKPKLGFPLIGETIQLKRLSEREPKRFQKDRKLPQVMTSLIGPTVL
CGAGNVLVSNEDKLVEMEGPKSFMLIGEDSVARQGDHRLRLARFLQAOL
QNTLRSSSEIGHFNEKMGKDEKVLPLVRLISLSTLFDVNDGHQOKLHL
LETLVGLSLVPLDFPTGRYRKGLQAKLDELISLIRRRDLRSGIASDDLLS
VLLTFREKNSLIDQIILNFSAMFASVDTYAPALIFKILYSPEYHEKVFQD
LETINKKESEELSMKDKSMKTYMCAVESLRYMPVRFKRAITDIHDGTTIRK
GMRLSPYTHIRREERPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPE
IETLFFVHVFVKNFSSTIPVDPKELSDPLPLPANGFSIKLPPRS"

BASE COUNT      429 a      315 c      314 g      436 t
ORIGIN

alignment_scores:
      Quality: 496.50      Length: 498
      Ratio: 1.661      Gaps: 11
      Percent Similarity: 60.040      Percent Identity: 27.510

alignment_block:
US-09-668-482-32 x AF318211 ..

Align seg 1/1 to: AF318211 from: 1 to: 1494

4 ProAlaLeuAlaLeuAlaLeuCysThrPheValLeuProLeuLeu 20
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
61 COTGTATCTCTTCCCTTACCCTCGCACCTATCTCGCATATCTTCT 110
20 uPheLeuAlaLeuLysLeuThrPaspLeuTrcysValSerSerArg 37
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
111 CTGCTCTCCGTTAC.....A 127
37 sPArgSerCysAlaLeuProLeuProGlyThrMetGlyPheProhe 53
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
128 ATCCACGATCCTCTGTAACTCCCTCGAAGATGAGTTTCTCTCTC 177
54 PheGlyGluThrLeuGluMetValLeuGlnArg..... 65
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
178 ATCGGGGAGACCATATATATGCGGACATCCCATCAGAAACCTCA 227
66 .LysPheLeuGluMetLysArgGlyLysTyrGlyPheIleTyrLysThr 82
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
228 AAAGTTTGTATGATGATGAGAAATTCGTCCTTACATGACTT 277
82 iLeuPheGlyArgProThrValArgValMetGlyAlaAspAsnValArg 98
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
278 CCTTAATGGGATCCACAGTTGTACTCTCGGGCTCGCGGAAACAA 327
99 ArgIleLeuGluGlyLysIleArgLeuValSerValHisTrpProAla 115
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
328 TTACTTCTTGGAAAGAGACAGCTGTAGAGATGAGAAAGGCCCAATC 377
115 rValArgThrIleLeuGlyAlaGlyCysLeuSerAsnLeuHisAspSer 132
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
378 TTTCATGAAACTGATGGGAAAGATTCATTTGCTAAAGAGCGGAG 427
132 eHisLysGlnArgLysValIleMetGlnAlaPheSerArgGluAla 148
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
428 ATCTTCGATCTTACGACATGCTGCTGCTTTGGGCGCTCAACT 477
149 LeuGlnCysTyrValLeuValIleAlaGluGluValSerSerCysLeu 164
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
478 TTCAAAATATTTCTGGGAGATGATTCAGAAATAGGACACCATTTCAA 527
165 .GluGlnTrpLeuSerCysGlyLysArgLysLeuLeuValTyrProGlu 181
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
528 TGAATAATGAG.....GGTAAAGATGAGTGAAGTGCCTCTTGG 571
181 alLysArgLeuMetPheArgIleAlaMetArgIleLeuGluCysGlu 197
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
572 TAAAGAGGCTTATCTCTCAATGCAAGACCCGTTTTTC..... 612

```

```

198 ProGlyProAlaGlyGlyGluAspGluGlnGlnLeuValGluAlaPh 214
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
613 .....GATGTAATGATGACCAACAGAACCACTTCATCATCTTTCT 656
214 eGluGluMetThrArgAsnLeuPheSerLeuProIleAspValProPhe 231
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
657 GGAACATATCTCTGGGAGAAAGTTGTCACTCCCGCGGACTTCCAGAA 706
231 eGlyLeuTyrArgGlyValLysAlaArgAsnLeuIleHisAlaArgIle 247
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
707 CTGCTTATCGTAAAGGCTTCAGCGCGG.....CTGAAGCTT 744
248 GluGluAsnIleArgAlaLysIleArgArgLeuGlnAlaThrGluPro 264
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
745 GATGAATTTCTCTCTCTTAATAAACGACAGAGAGATGCTGCGCTTC 794
264 pGly.....GlyCysLysAspAlaLeuGlnLeuLeuIleGluHis 278
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
795 AGCAGATAGCTTCTGATGATGACAGATCTACTGCGGTGCTCACCTTCA 844
278 eTrpGluArgGlyLysArgLeuAspMetGlnAlaLeuLysGlnSerSer 294
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
845 GAGATGAAAGGAACTCCTACAGACAGACAGGAGATTCTGACAACTTT 894
295 ThrGluLeuPheGlyGlyLysGluThrThrAlaSerAlaAlaThrSe 311
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
895 TCTGCTATGTTTCATGCTTCATATGACACACTGTTCGACCAATGGCTT 944
311 rLeuIleThrTyrLeuGlyLeuTyrProHisValLeuGlnLysValArg 328
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
945 GATATTTAAGCTTCTATACTCCATCTCGAATACCATGAAAGATTTTC 994
328 LuGluIleLysSerLysGlyLeuLeuCysLysSerAsnGlnAspAsnLys 344
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
995 AAGAG.....CAGTTGAAATATTTGGCAATYAAAAGAGAGAGAGAA 1038
345 LeuAspMetGluThrLeuGluGlnLeuLysTyrIleGlyCysValIle 361
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1039 ATCAGTTGGAAGATTTGAATCTATGAATATATACATGCAAGCATTTCA 1088
361 sGluThrLeuArgLeuAsnProProValProGlyLysPheArgValAla 378
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1089 AGAATCACTCAAGATGATGCCACAGTTTGGATATATTTCTGAAGCTTA 1138
378 eLysThrPheGluLeuAsnGlyTyrGlnIleProLysGlyTyrAsnVal 394
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1139 TCAGTATATTCATATATGATGGTATACAAATTCAAAAGATGAGGCTT 1188
395 IleTyrSerIleCysAspThrHisAspValAlaAspIlePheThrAsn 411
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1189 TTATTTACCTTTACTATACATCTGAGAGAGAGTACTTCCCTGAGCC 1238
411 sGluGluPheAsnProAspArgPheIleValProHisProGluAspAla 428
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1239 TGAAGATTTCAAGGCTTCAAGATT.....GAGATGAG 1273
428 eArg.....PheSerPheIleProPheGlyGlyLeuArg 440
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1274 GCAGGATGATGACTCTTACACATATGTACCATTTGGAGGAGGCTCGCG 1323
441 SerCysValGlyLysGluPheAlaLysIleLeuLeuLysIlePheThr 457
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1324 ACATGCTCCAGATGGAAATTTTCAAGATTAAGATTAATGCTTTGTCCA 1373
457 GluLeuAlaArgHisCysAspTrpGlnLeuLeuAsnGlyProProThr 474
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1374 TCATTTGCTTAAATATTCACAGAGCTACATTCAGTTATCCCATGAAA 1423
474 eLysThrSerProThrValTyrProValAspAsnLeuProAla 488
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1424 AAGTTTATTCAGAT.....CCACTACCTCTCTCCCTGCG 1458

```

seq_name:	gb_pl1:AC006931	FEATURES	Location/Qualifiers
seq_documentation_block:		source	1. 116205
LOCUS	AC006931 116205 bp DNA		/organism="Arabidopsis thaliana"
DEFINITION	Arabidopsis thaliana chromosome II section 231 of 255 of the complete sequence. Sequence from clones F14N22, F7D19.		/cultivar="Columbia"
ACCESSION	AC006931 AE002093		/db_xref="taxon:3702"
VERSION	AC006931.5 GI:6598641		/chromosome="II"
KEYWORDS	HTG.		complement(1..2131)
SOURCE	thale cress.	misc_feature	/note="Sequence from clone F14N22"
ORGANISM	Arabidopsis thaliana	mRNA	<1226..>2344
	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.	gene	/gene="At2g42650"
REFERENCE	1 (bases 1 to 116205)		<1226..>2344
AUTHORS	Lin, X., Kaul, S., Rounsley, S.D., Shea, T.P., Benito, M.-I., Town, C.D., Feldblyum, T.V., Buel, C.R., Ketchum, K.A., Lee, J.J., Ronning, C.M., Koo, H., Moffatt, K.S., Cronin, L.A., Shen, M., Vanaken, S.E., Umayam, L., Mallon, L.J., Gill, J.E., Adams, M.D., Carrera, A.J., Creasy, T.H., Goodman, H.M., Somerville, C.R., Coppenhaver, G.P., Preuss, D., Nierman, W.C., White, O., Eisen, J.A., Salzberg, S.L., Fraser, C.M. and Venter, J.C.	CDS	/note="At2g42650"
TITLE	Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana		/gene="At2g42650"
JOURNAL	Nature 402 (6763), 761-768 (1999)		/note="hypothetical protein"
MEDLINE	20083487	misc_feature	/codon_start=1
PUBMED	10617197	mRNA	/protein_id="AAF18664.1"
REFERENCE	2 (bases 1 to 116205)		/db_xref="GI:6598642"
AUTHORS	Lin, X.	gene	/translation="MSRYSPTVDVAVALVKEGNEKSRTEKPOLLEEDGPFYLVAL/KTIQRMNTNMYRPIPLPPLINTEDSPCLLITDDREGSLTEEDAKKNSENIPIT/KVVKLSLKSDYSSFEKRLCDSYMFSDRVTIPMLPTLICKKFPQSKKTPAID/LKHNWKEQIERKACGAAPFMRGSCSAIKVAKLSMESDDIVENTATLAKVDVPLS/RMKYIRSLHLKLSLSLPLIYQYPTQLKIDPLGVEVRKNGEGLAKSDVDDSSKSV/KTKKNGKIHEDRYWDSNVSELTIGDEDRSDRVEDVADLNASGDKRRKRMSSKS/AVSGKPDIVKSNKGSKKLKIDDESGGFRKATKR"
JOURNAL	Direct Submission		complement(2132..>116205)
COMMENT	Submitted (09-MAR-2000) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA		/note="Sequence from clone F7D19"
	On Dec 17, 1999 this sequence version replaced gi:4512656.		complement(join(<2761..2943,3035..3164,3258..3372,3450..3489,3570..3646,3732..>3954))
	The sequence and annotation of chromosome 2 were merged from those of the individual clones on this chromosome after removing overlaps. For detailed information, please see the TIGR web site (http://www.tigr.org/tdb/at/at.html).	gene	/gene="At2g42660"
			complement(<2761..>3954)
		CDS	/note="F7D19.34; predicted by gensecan"
			complement(join(2761..2943,3035..3164,3258..3372,3450..3489,3570..3646,3732..3954))
			/gene="At2g42660"
			/note="hypothetical protein"
			/codon_start=1
			/protein_id="AAd21740.1"
			/db_xref="GI:4512686"
			/translation="MAGRSPPGKCTWLDLNRETVENKNEESSVEDEDOITNNVRO/YIRSNMPLRMTPDLHLSEFVAVRATGPDRTPLVLEMMNLKGLSLAHVSHQMY/RSKRLEPSRPGFGAFMSGORSYVLRMDMDISRCIPSDLRHAYNSKTVSPYADQAVY/TNLGNTFLMRSSMSFGICLRNDRDSTKTKTLPLEIRKKNKKRYARDEVSYSKRLKS/MSGGIIQPEFGNCOKRPTDITNMLSLSTSSP"
			3681..3724
		repeat_region	/rpt_family="(GA)n"
			4087..4204
		repeat_region	/rpt_family="(7A)n"
			5736..5806
		repeat_region	/rpt_family="(GA)n"
			complement(join(<7700..7818,7906..8112,8195..8348,8844..8989,9294..>9396))
		mRNA	/gene="At2g42670"
			complement(<7700..>9396)
		gene	/gene="At2g42670"
			/note="F7D19.33; predicted by gensecan"
			complement(join(7700..7818,7906..8112,8195..8348,8844..8989,9294..9396))
		CDS	/gene="At2g42670"
			/note="hypothetical protein"
			/codon_start=1
			/protein_id="AAd21739.1"
			/db_xref="GI:4512685"
			/translation="MPYFQRLYNCTCKASFSSDGPITTEDALEKRVNVLEKIRPSDVGI/BDQADLRSGPLNERNOSPAIKYLHLHEDSFGIGFCMPSSMILNHPG/MVLISKVIGSMHVSVDMLPOLTEPDSQEAQPAKVKVTEMTAOSPVTLPKPS/GGNTRCFKATICALDILAIAPYSSSEHRRHCTYFRKSRREDLPGELNDGEVYVDTW/LIEFPDPDEVIRIRIPYRGVYIRT"
			10075..10153
		repeat_region	/rpt_family="(GAA)n"
			join(10610..10809,11165..11248,11522..11604,11693..11994)
		mRNA	

We thank the CSHL/WashU/ARI consortium for sequencing BAC clones F6P23, F5U6, T17A5, and T13L16, the ESSA group for sequencing clone F13D4, and Scott Jackson, Jiming Jiang, Klaus Meyer, Eric Richards and Satoshi Tabata for helpful assistance. In addition, we would like to thank the TIGR Bioinformatics Department, especially Likun Zhou, Hanif Khalak, Michael E. Heaney, Lily Fu, Feng Liang, Jeremy Peterson, Michael Holmes, and Delwood Richardson for software and database support.

This work was supported by the National Science Foundation, Department of Energy and the US Department of Agriculture.

Address all correspondence to: at@tigr.org.

```

gene      /gene="At2g42680"
          /10610..11994
          /gene="At2g42680"
          /note="F7D19.32"
          /join(10668..10809,11165..11248,11522..11604,11693..11812)
          /gene="At2g42680"
          /note="unknown protein"
          /codon_start=1
          /protein_id="AAD21738.1"
          /db_xref="GI:4512684"
          /translation="MAGIGPIPTDMEPVYIRKKPANAAKDEKTVNAARRSGADIEF
VRKFNAGTNNKAASGTSILNTKMLDDPTENLTHERVPELTKAIQWARTDKKLTQSOLA
QIINEKPOVIOEESGKAIPNOQILISKLERALGKLRGK"
          /complement(join(11994..12273,12362..>13378))
          /gene="At2g42690"
          /complement(11994..>13378)
          /note="At2g42690"
          /complement(join(12052..12273,12362..13378))
          /gene="At2g42690"
          /codon_start=1
          /product="putative lipase"
          /protein_id="AAD21737.1"
          /db_xref="GI:4512683"
          /translation="MATTTTSWELLGSKNMDTILDPDLSRELILRCDFQCATYD
AFVNDNSKYGCGASRYGKSFVKMLENASDEYVNFYATARSLPEGLDLOSOR
DSWDRESNMFGYIAVTSDESKALGRREIYIALRGSTRNEMVAVLGAAPTSDPLH
GPDODSGGVYEGEPTPDSDEEGCGVMGLTIYSNHPESKFTLSRSQILAKI
KELLKYDKPKPSIVTLGHSLGATEANLAYDAENGSDVPEYTAIVFCPOVGNKE
FREDVSHKMKILIHVANTIDLRTIPGGILGYVDIDINVIDIRKSPISDSNPGD
WHNLQAMLHVAVMNGCKGFEKLMVKRSIALVYKSCFEKLAECILVPSGVMWEKKNGLI
KNDEGMVLAIVVEEVPYEF"
          /join(<14160..14207,14422..14565,14779..14890,15255..15434,
15513..15664,15832..17364,17446..17514,17602..17786)
          /gene="At2g42700"
          /note="At2g42700"
          /note="F7D19.30"
          /join(14160..14207,14422..14565,14779..14890,15255..15434,
15513..15664,15832..17364,17446..17514,17602..17730)
          /gene="At2g42700"
          /note="unknown protein"
          /codon_start=1
          /protein_id="AAD21736.1"
          /db_xref="GI:4512682"
          /translation="MALIDVAISCLNSIREIEDVKDAIYIYDAGCTESFOFAPFL
FELGARAVCSLENNMTSLDADVADNMSKDCAKRIVMTSLNDAAHYMTRCISTHGC
PDAYREYETILVDYNEHTKSKDKGVSKFSSALBESLIMEPISENVDISGGA
QGSVAASISROHEDLSREGLPISTGSMKSDTDDVPGATLTHFLYQLAKMELKLE
IISLDGDSKSNVGLITDMSVYDVARKRSAGLLVDRTDLITPCGHPGSLDRIFS
SLPRARFSSQAQLKQGVPSINRPSLDVQVPLGELINEEFSKTRDSCGLPGISAFLEFG
WISYTSAPQNVGLFNECDKSKTNWTELLNGSLVATCEFGFTYLEAMIDRKDGAV
LVKMKLOEALRENIENVNARPGVATKPELOAMIKRALSQSSSLKNKGIITOLGAT
AAALDESQAKMDTFSSAEMMLNVASGDTSGIAADISDILNSAYAELOAKNNEKD
SSRGILSRDALLITVIYIILAGENPFTSGSGPFSWDEEHLTKGIVDALENBSA
GNIKPLNGLEELERLNRKSEETKIPSDDLDDIDALDDEWKGWDEEVEVNS
KADESYDDMLKIDLNRDVSLEFRFLHKLISLRTNRLPREGSIASBSPGPGSGNK

```

alignment_scores:

Quality:	460.00	Length:	526
Ratio:	1.655	Gaps:	14
Percent Similarity:	52.852	Percent Identity:	26.046

alignment_block:

US-09-668-482-32 x AC006931 ..

Align seg 1/1 to: AC006931 from: 1 to: 116205

```

44 LeuProPoglyThrMetGlyPheProPhePheGlyIuThrLeuGlnMe 60
   ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
67490 CTTCTCTCGTGAAGAAATGGGCTTCATGATGTGTGAACAATGGAATT 67539
   ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
60 tValLeuGlnArgArgLys.....PheLeuGlnMetL 71

```

```

67540 CTACAAAGCTCAGAAAAGCAACAGAGCTTTTGGAGATTTTGTAAATCCAA 67589
      :               :|||:
71 ysArGArgLysTyrglyPheIleTyrgLysThrHisLeuPheGlyArgPro 87
   |||:||||| |||:|||||:|||||:|||||:||||| |||
67590 GGATCATCAACAAAGTGGAGCAATCTTCAAGCAAGATCATGGGATGACCA 67639
      :|||:|||||:|||||:|||||:|||||:|||||:|||||
88 ThrValArgValMetGlyAlaAspAsnValArgArgIleLeuGlnGly 104
   |||:||||| |||:|||||:|||||:|||||:||||| |||
67640 ACAATGATGATCAATGAGACTGAGGCAATAGCTCATATATACCAACAA 67689
      :|||:|||||:|||||:|||||:|||||:|||||:|||||
104 uHisArgLeuValSerValHisIleProAlaSerValArgThrIleLeuG 121
   |||:||||| |||:|||||:|||||:|||||:||||| |||
67690 GTTACTCTGGGTGAGTACGCTCGTGGCTTCATCTCTGTTCACACTATGG 67739
      :|||:|||||:|||||:|||||:|||||:|||||:|||||
121 lValAlaGlyCysLeuSerAsnLeuHisAspSerSerHisLysGlnArgLys 137
   |||:||||| |||:|||||:|||||:|||||:||||| |||
67740 GCATGAACATGCATCATGCTTAACAAAGCGAGAGCATCGGTCCTCAGA 67789
      :|||:|||||:|||||:|||||:|||||:|||||:|||||
138 LysValIleMetGlnAlaPheSerArgLysIleLeuGlnCysTyValLe 154
   |||:||||| |||:|||||:|||||:|||||:||||| |||
67790 GGGATTTGATGTAATATGCTTGTAGCTACATTTGGTGAAGTCTTAATACC 67839
      :|||:|||||:|||||:|||||:|||||:|||||:|||||
154 uValIleAlaGluGluValSerSerCysLeuGlu...GlnTrpLeuSerC 170
   |||:||||| |||:|||||:|||||:|||||:||||| |||
67840 TAAGCTCTGTGATACCGCTTAAGTTTCATCATGAAACCGAATGGCGCA... 67885
      :|||:|||||:|||||:|||||:|||||:|||||:|||||
170 yGlgGlyArgGlyLeuLeuValTyrrProGluValLysArgLeuMetPhe 186
   |||:||||| |||:|||||:|||||:|||||:||||| |||
67886 ..GGCAAGGAAGAGATTTACCTTCCTACCGTTCGCAAAAGTCTCACTTTC 67933
      :|||:|||||:|||||:|||||:|||||:|||||:|||||
187 ArgIleAlaMetArgIleLeuLeuGlnCysGluProGlyProAlaGly 203
   |||:||||| |||:|||||:|||||:|||||:||||| |||
67934 ACTGTATGATTTTGAAGCTTTGTATGGATTAAGAGC..... 67969
      :|||:|||||:|||||:|||||:|||||:|||||:|||||
203 yGlyGluAspGluGlnGluLeuValGluAlaPheGluGluMetThrArgA 220
   |||:||||| |||:|||||:|||||:|||||:||||| |||
67970 .....GAGATTGGATGTTAGAGGCTTTGAGAGGTTTAAAGAG 68009
      :|||:|||||:|||||:|||||:|||||:|||||:|||||
220 snLeuPheSerLeuProIleAspValProPheSerGlyLeuTyrrArgGly 236
   |||:||||| |||:|||||:|||||:|||||:||||| |||
68010 GGGTTTGCCTTGCCTGCGGTTGAGTTCTTGTCTAAGTTCTCTAGAGCA 68059
      :|||:|||||:|||||:|||||:|||||:|||||:|||||
237 ValLysAlaArgAsnLeuIleHisAlaArgIleGluGluAsnIleArgAl 253
   |||:||||| |||:|||||:|||||:|||||:||||| |||
68060 AAGAAAGCAAGCTAGAGATTAAGACATTTCTAGTTGGAAAGCTACGGA 68109
      :|||:|||||:|||||:|||||:|||||:|||||:|||||
253 aLysIleArgArgLeuGlnAlaIleThrGluProAspGlyCysLysAspA 270
   |||:||||| |||:|||||:|||||:|||||:||||| |||
68110 GAAGGAGAGAGAAATG.....GAAAAGAGGAGAGCTGAGAACCCAA 68150
      :|||:|||||:|||||:|||||:|||||:|||||:|||||
270 lalLeuGlnLeuLeuIleGluHisSerTrpGluArgGlyGluArgLeuAsp 286
   |||:||||| |||:|||||:|||||:|||||:||||| |||
68151 ACACAAAGCTTTTC.....TCGAGGCTAGTA 68176
      :|||:|||||:|||||:|||||:|||||:|||||:|||||
287 MetGlnAlaLeuLysGlnSerSerThrGlu..... 296
   |||:||||| |||:|||||:|||||:|||||:||||| |||
68177 GAAGAGTTGATCAAGAGTGATACCGAAGAGAGGTGTGATATAATAT 68226
      :|||:|||||:|||||:|||||:|||||:|||||:|||||
297 .....LeuLeuPheGlyGlyHisGluThrThrAlaSerAlaIleArg 311
   |||:||||| |||:|||||:|||||:|||||:||||| |||
68227 GGTTTTGCTAGTGTTCGAGCTCATGACACAAACCTTACGCCATGTCCA 68276
      :|||:|||||:|||||:|||||:|||||:|||||:|||||
311 eTrLeuIleThrTyrrLeuGlyLeuTyrrPro..... 320
   |||:||||| |||:|||||:|||||:|||||:||||| |||
68277 TGACTTTCAGATGTTAGCTGACGACCCCAACTGTGCGACACTCTCCT 68326
      :|||:|||||:|||||:|||||:|||||:|||||:|||||
321 .....HisValLeuGlnLys 325
   |||:||||| |||:|||||:|||||:|||||:||||| |||
68327 CAAGGTAAACATCAACATGATATATGATGTCTCTCTCTCAACCAAAA 68376
      :|||:|||||:|||||:|||||:|||||:|||||:|||||
325 sValArgGluGluIleLys.....SerLysGlyLeuL 336
   |||:||||| |||:|||||:|||||:|||||:||||| |||

```

```

68377 ATCATATCTCAACTCTCAGTTTCTCTTGATTCATTCAGAACTGCTC 68426
336 euCysLysSerAsn.....GlnAspAsnLysLeuAspMetGluThrLeu 350
||||:|||||
68427 AATATAAGCAACAAGGAGGAGCAATATCTTACAGTGGAAAGATGTG 68476
351 GluGlnLeuLysTyrIleGlyCysValIleLysGluThrLeuArgLeuAs 367
||||:|||||
68477 AAAAAGATGAGATAGTAGTGGCAAGTTGTTCGTAACAACTCGATTATC 68526
367 nProProValPProGlyGlyPheArgValAlaLeuLysThrPheGluLeuA 384
||||:|||||
68527 CCTCCAACTCTTCGGTTCCTTTCAGAAAGCAGTGGCTGATATTGACTATG 68576
384 snGlyTyrGlnIleProLysGlyTyrPasnVal..... 394
||||:|||||
68577 GAGGATACACATCCCAAGAGCTGGAAAGTTTAACACATCCATCTCTTA 68626
395 ..... 395
68627 AGATTCTCAGCTTCTATCAAAACTCTATTCTTTGAAAGAAATACTAAT 68676
395 e.TyrSerIleCys..... 399
||||:|||||
68677 AATTAACCTCTGCTCTTCTTTGTTGATGCACATCTCTGGACCACTTAT 68726
400 AspThrHisAspValAlaAspIlePheThrAsnLysGluGluPheAsnPr 416
||||:|||||
68727 GGAACATATTATATCCGTGAATATTTCAGAAATCCAAATGATTTTGATCC 68776
416 nAspArgPheIleValProHisProGluAspAlaSerArgPheSerPheI 433
||||:|||||
68777 AACAAATTTTGATTAACCA.....ATACAGGCATATACATATC 68814
433 LeuProPheGlyGlyGlyLeuArgSerCysValGlyLysGluPheAlaLys 449
||||:|||||
68815 TACCATTTTGGAGGTGGACCAAGCTTTGGCAGTGCACCAACTGACCAAA 68864
450 IleLeuLeuLysIlePheThrValGluLeuAlaArgHisCysAspTrpG 466
||||:|||||
68865 ATCAGCATCTCTGTTTCATGATTTGTTGTGACAGGTTTGTGACGTGC 68914
466 nLeuLeuAsnGlyProProThrMetLysThrSerProIleValTyrProV 483
||||:|||||
68915 CTTAGCTTACCCAGATGAGCAATTAGCATGATCCTTCCCATTTCCGT 68964
483 alAspAsnLeuProAlaArgPheThr 491
||||:|||||
68965 CCCTTGGAAATGCCAATCAAAATTCT 68990
seq_name: gb_pat1:AR074363
seq_documentation_block:
LOCUS AR074363 1608 bp DNA PAT 28-AUG-2000
DEFINITION Sequence 1 from patent US 5952545.
ACCESSION AR074363
VERSION AR074363.1 GI:10001118
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 (bases 1 to 1608)
Koncz,C., Matnir,J., Szekeres,M. and Altmann,T.
Nucleic acid molecules encoding cytochrome P450-type proteins
involved in the brassinosteroid synthesis in plants
Patent: US 5952545-A 1 14-SEP-1999;
JOURNAL
FEATURES
source
1..1608
/organism="unknown"
BASE COUNT 382 a 374 c 374 g 478 t
ORIGIN

```

```

alignment_scores:
Quality: 450.00 Length: 503
Ratio: 1.562 Gaps: 15
Percent Similarity: 57.256 Percent Identity: 26.640
alignment_block:
US-09-668-482-32 x AR074363 ..
Align seg 1/1 to: AR074363 from: 1 to: 1608
6 LeuLeuAlaSerAlaLeuCysThrPheValLeuProLeuLeuPheLe 22
||||:|||||
72 CTCCTCTTTCATCGCCGCCGCCGCTTCTCTCTCTACTC..... 110
22 uAlaAlaLeuLysLeuThrPheLeuTyrCysValSerSerArgAspArg 39
||||:|||||
111 .....CGCGCTACACGTT 123
39 erCysAlaLeuProLeuProGlyThrMetGlyPheProPheGly 55
||||:|||||
124 ACCGTGGATGGGTGCTGCTCGGAGACCTTGCTCTCTCATATAGA 173
56 GluThrLeuGlnMetVal.....LeuGlnArgArgLysPh 67
||||:|||||
174 GAGACTTTTCAAGCTGATCGAGCTTACAAACAGAAACCTGAGCCTTT 223
67 eLeuGlnMetLysArgArgLysTyrGlyPheIleTyrLysThrHisLeu 84
||||:|||||
224 CATGCACAGAGAGATGACCCGCTGACGTTGCGTTTCTCATGACGATCTT 273
84 heGlyArgProThrValArgValMetGlyAlaAspAsnValArgArgIle 100
||||:|||||
274 TTGCTGAACCCGACGATTTTTCACAGTACCCGGAACAGACCGTTTGT 323
101 LeuLeuGlyGlnHisArgLeuValSerValHisTrpProAlaSerValAr 117
||||:|||||
324 CTTCAGAACGAAAGGAGACCTTTTGTGATGTTCTTATCCTGCTTCATTTG 373
117 gThrIleLeuGlyAlaGlyCysLeuSerAsnLeuHisAspSerHisSL 134
||||:|||||
374 TACCTTTTGGGAAACACTCTCTCTCTTATGAAAGTTCTTTGCTTA 423
134 ysGlnArgLysLysValIleMetGlnAlaPheSerArgGluAlaLeuGln 150
||||:|||||
424 AACGATATGACCTCTCTCAACATGACCTTGTCAATTTCTCAATATAA 473
151 CysTyrValLeuVal...IleAlaGluGluValSerSerCysLeuGluG 166
||||:|||||
474 GACCATCTCATGCTTGATATGACCGGTTAGTCGCTTAAATCTTGATTC 523
166 nThrPheLeuSerCysGlyGluArgGlyLeuLeuValTyrProGluValLys 183
||||:|||||
524 TTGCTCTCT.....CTGTTCTCTCATGGAAGAAAGCAAAA 561
183 rGluMetPheArgIleAlaMetArgIleLeuLeuGlyCysGluProGly 199
||||:|||||
562 AGATTAACGTTTGAGCTAACGCGAAGCACTGATGAGCTTATATCA... 608
200 ProAlaGlyGlyGlyLysAspGluGlnGlnLeuValGluAlaPheGluG 216
||||:|||||
609 .....GGGGAATGAGAGTGAAGTTTAAGAAAGATATCTTCT 646
216 uMetThrArgAsnLeuPheSerLeuProLeuAspValProPheSerGlyL 233
||||:|||||
647 TGCATCGAAGGCTTCTTCTCTCTCTCTCTCTC...TTCTCCACA 693
233 euTyrArg...GlyValLysAlaArgAsnLeuIleHisAlaArgIleGlu 248
||||:|||||
694 CTACCGCAAAAGCATCAAGCGCGAGG.....AAGTGCGG 731
249 GluAsnIleArgAlaLysIleArgArgLeuGlnAlaThrGluProAspG 265
||||:|||||
732 GAGCGTTGACGCTGTGTGATGAAGAAAGAGGAGGAGGAGAAAGG 781

```

```

265 yely.....CysLysAspAlaLeuGlnLeuLeu.....GluHis 278
      ||||| ||| |||||
782 AGCGAGAGAAAGAAAGATATGCTTGGCGGCTTGGCGGCGGAGATG 831
      |||||
278 eTTrGluArgGlyGluArgLeuAspMetGlnAlaLeuGlnSer 294
      ||| |||||
832 GATTTTCGATGAGAGATTTGTGACTTC.....TTC 863
295 ThrGluLeuPheGlyGlnGluThrPheAlaSerAlaAlaThr 311
      ||||| ||||| ||||| ||||| ||||| |||||
864 GTGGCTTACTTGTGCGGTTATGAAACACCTCCAGCATGATGATCT 913
311 rLeuLThrTyrlLeuGlyLeuTyrrProHisValLeuGlnValArg 328
      ||||| ||||| ||||| ||||| ||||| |||||
914 CGCGCTCAAAATTTCTCAGCAGACCTCTTACCTTGTGTCACACCAAG 963
328 LuGluLLeuLysSerLysGlyLeuGlySerAsnGlnAspAsnLys 344
      ||||| ||||| ||||| ||||| ||||| |||||
964 AAGAG.....CATGAAAAGATTAGGCAATGAGAGATTCGTATAGT 1007
345 LeuAspMetGluThrLeuGlnLeuLysTyrlLeuGlyValLLeu 361
      ||||| ||||| ||||| ||||| ||||| |||||
1008 CTGATATGATGATTTACAGTCATGCCATTTCACACATGTGTGTTAA 1057
361 sGluThrLeuArgLeuAsnProProValProGlyLysPheArgValAla 378
      ||||| ||||| ||||| ||||| ||||| |||||
1058 TGAGACGCTAGAGAGCTACATCATCGCGGCTTTTCAGACGTGCA 1107
378 eLysThrPheGlnLeuAsnGlyTyrlGlnLLeuProLysGlyTrpAsnVal 394
      ||||| ||||| ||||| ||||| ||||| |||||
1108 TGACGAGATGTTAGAGTCAAGGTTATAAATTCCAAAAGGTTGGAAGTA 1157
395 lLeuTyrrSerLysAspThrHisAspValAlaAspLLeuPheThrAsnLys 411
      ||||| ||||| ||||| ||||| ||||| |||||
1158 TTCATCGCTTAGAGCGGCTTCATTAGACCCAAACCACTTCAAGATGC 1207
411 sGluGluPheAsnProAspArgPhe.....lLeuValProHisP 424
      ||||| ||||| ||||| ||||| ||||| |||||
1208 TCGACATTTCAACCTTGAGATGCGAGACAGTCCGTTACGACAGCGC 1257
424 rGluGlnAspAlaSerArgPheSerPheLLeuProPheGlyGlyLeuArg 440
      ||||| ||||| ||||| ||||| ||||| |||||
1258 CTTCATATGTC.....TTCACACCGGTTGGGGGAGCGCAAGG 1295
441 SerCysValGlyLysGluPheAlaLysLLeuLeuLysLLeuPheThrVa 457
      ||||| ||||| ||||| ||||| ||||| |||||
1296 CTATGTCGCGGTAGAGCTGAGGTTGCTCTCTGTTTCTTCA 1345
457 lGluLeuAlaArgHisCysAspTrpGlnLeuLeuAsnGlyProProThrM 474
      ||||| ||||| ||||| ||||| ||||| |||||
1346 CGCGCTAGTGCAGGCTTCAGTTGG.....GTTCTCTGCGAG 1380
474 eLysThrSerProThrValTyrrProValAspAsnLeuProAlaArgPhe 490
      ||||| ||||| ||||| ||||| ||||| |||||
1381 AGCAAGACAGAGCTGTTTCTTCCAACTACAAGACGAGCAACGCTAC 1430
491 ThrTyrrPhe 493
      |||
1431 CGGATCTTC 1439
seq_name: gb_p12:ATCYP450R
seq_documentation_block:
LOCUS ATCYP450R 1608 bp mRNA PLN 13-MAY-1996
DEFINITION A.thaliana mRNA for cytochrome P450.
ACCESSION X87367
VERSION X87367.1 GI:853718
KEYWORDS Cytochrome P450.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosid 11;

```

```

REFERENCE 1 (bases 1 to 1608)
AUTHORS Szekeres,M., Nemeth,K., Koncz-Kalman,Z., Mathur,J., Kauschmann,A.,
Altmann,T., Redei,G.P., Nagy,F., Scheil,J. and Koncz,C.
TITLE Brassinosteroids rescue the deficiency of CYP90, a cytochrome P450,
controlling cell elongation and de-etiolation in Arabidopsis
JOURNAL Cell 85 (2), 171-182 (1996)
MEDLINE 96200769
REFERENCE 2 (bases 1 to 1608)
AUTHORS Szekeres,M.
TITLE Direct Submission
JOURNAL Submitted (22-MAY-1995) M. Szekeres, Inst. of Plant Biology,
Biological Research Center, Hungarian Academy of Sciences, PO Box
521, 6701 Szeged, HUNGARY
FEATURES
source Location/Qualifiers
1..1608
/organism="Arabidopsis thaliana"
/cultivar="columbia"
/db_xref="taxon:3702"
/chromosome="5"
/clone_1lb="lambda gt10"
/clone="C204"
/map="14.3 cm"
48..1466
/gene="CPD"
48..1466
/gene="CPD"
/function="loss of gene function results in dwarf
phenotype"
/note="cytochrome P450"
/codon_start=1
/product="CYP90 protein"
/protein_id="CA60793.1"
/db_xref="GI:853719"
/db_xref="SWISS-PROT:Q42569"
/translacion="MAFALDLSSIAAGFLILRLRRTRYRMGLPPGSLGILPILGET
FOLIGAYKTEPEPFIDERVARYGVFTHLEFPTIFSADPTNRFVLQNEKLEFC
SYPAICNLGKSLILMKGLSRHMSLSFANSIKDHLMDIDLRLVRLNLSM
SSRVILMEERKKTFFELTYKOLSPDGSMSLSREYLLIEGPRSLPLESTYR
KATQARRKVAELTYVMKRREBEERKADMLAALADDSDEETVDFVALV
VAGETTTSTIMLAVAFLETFPLAQLQLEBEKRTAMKSDSYLSMSDYKSKSPFQC
VNETLRVANIIIGGVFRAMTDVEIKGYRIKGMKVSSFRVAHLDPNHFKARFNP
WRMOSVTTGDSNFTPEGGGRLCPGYELARVALSVPLHRLVTFSWVPAEODKLV
FPTPTQKRYPLFVVRDFAT"
BASE COUNT 382 a 374 c 374 g 478 t
ORIGIN
alignment_scores:
quality: 450.00 length: 503
Ratio: 1.562 Gaps: 15
Percent Similarity: 57.256 Percent Identity: 26.640
alignment_block:
US-09-668-482-32 x ATCYP450R ..
Align seg 1/1 to: ATCYP450R from: 1 to: 1608
6 LeuLeuAlaSerAlaLeuGlyThrPheValLeuProLeuLeuPheLe 22
|||||:|||||
72 CTCCTCTCTTCATCGCGCGCGGCTTCTCTCTCTACTC..... 110
22 uAlaAlaLeuLysLeuThrPAspLeuTyrrCysValSerSerArgAspArg 39
|||||
111 .....CGCGCTACAGCTT 123
39 eCysAlaLeuProLeuProProGlyThrMetGlyPheProPhePheGly 55
::: ||||| ||||| ||||| ||||| ||||| |||||
124 ACCGTCGATGGGCTCTCCCGGGAAGCCTTGCTCCCTGATAGAGA 173
56 GluThrLeuGlnMetVal.....LeuGlnArgAlaGlySph 67
|||||:|||||:|||||
174 GAGACTTTTCAAGCTGATCGAGCTTACAAAACAGAGAACCTTGAGCTTT 223

```



```

/protein_id="AAB17070.1"
/db_xref="GI:1421741"
/translation="MAFFILFISFGFICFTALLRMNQYKNOKNLPGTGMPLF
GETTEFLKIPSEFKKNORARYSEFKSHILGCGTIVMDSLNRYIIIVNEKGLVPEY
POSMIDILGKNIAAANGSAHKYRGALLISLSPTRDQLPRIDEMRSHLTNWDN
KVIDIQRTKMAFLSKQJAGLESTSLAOFSESEFNVLGLSLP INLPNTYR
GFOARKIIVMLRTLIERERASKELIQDMGLYMEARTRKLDDEMDLITILVS
GYETVSTSMMAVYLIHDHPKVLIELREHMLAIEKKRPEDPIDYNDYRSMTFAVI
LETSLATIVNGVLRTQDMENINGYIIIPKGMRYVYTRRLNDPRLYDPDPSNPMR
MMDKSLIEHONSPLVFGGSTRQCPGKEIGVAIEISFLHYFTKRYMEELGDKLMKPPR
VEAPNGRIRIVSAH"

```

```

BASE COUNT      470 a      242 c      268 g      415 t
ORIGIN

```

```

alignment_scores:
  Quality: 442.00      Length: 463
  Ratio: 1.637      Gaps: 10
  Percent Similarity: 58.315      Percent Identity: 26.998

```

alignment_block:

```
US-09-668-482-32 x SLU54770 ..
```

```
Align seg 1/1 to: SLU54770 from: 1 to: 1395
```

```

18 LeuLeuLeuPheLeuAlaAlaLeuLysLeuTrpAspLeuTyrCysValSe 34
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
40 CTAATGATATCTTTGTACTGCTCTTATTAAAGATGAATCAAGTCACAGATA 89
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
34 rSerArgSparSerCysAlaLeuProLeuProGlyThrMetGly 51
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
90 CCAAAAAAAC.....TTGCCCTGCTACTAGTGGGT 121
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
51 heProPheheGlyIuThrLeuGlnMetValLeuGlnArgArgysPhe 67
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
122 GCCCCTTTTGTGAACTACTGAGTTCTTAAACCTTGTCACAACTTTC 171
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
68 LeuGlnMetLysArgArgLysTyrGlyPheIleTyrLysThrHisLeu 84
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
172 ATGAAACCAAGAGCCAGATATGGAGCTTTTAAATCACACACTACT 221
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
84 eGlyArgProThrValArgValMetGlyAlaAspAsnValArgArgIle 101
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
222 TGGTGTCCCAACAATGTTCAATGATTCAGACCTGACACAGATATATAC 271
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
101 euLeuGlyGlnHisArgLeuValSerValHisTrpProAlaSerValArg 117
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
272 TAGTGAATGAAGCGAAGACTGTGCCAGATACCCACAGCTCTATGATA 321
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
118 ThrIleLeuGlyAlaGlyCysLeuSerAsnLeuHisAspSerSerHisLys 134
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
322 GATATTTAAGAAATGTAAATATTCAGCTGCATCATGTTCAAGTCACACA 371
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
134 sGlnArgLysLysValIleMetGlnAlaPheSerArgGluAlaLeuGlnC 151
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
372 GTACATGTGAGGGGTGATTCCTTAATAGCCCTACACATGATGACAG 421
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
151 ySlyValLeu...ValIleAlaGluGluValSerSerCysLeuGln 166
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
422 ACCAACTTTTGCTTAAATGTATGAGATTATGATCCACACTTAACCAAT 471
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
167 TrpLeuSerCysGlyIuArgGlyLeuLeuValTyrProGluValLysArg 183
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
472 TGG.....GATAATAAGTATTATGACATTCAAGAGAAACCAATTA 512
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
183 gluMetPheArgIleAlaMetArgIleLeuLeuGlyCysGluProGlyP 200
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
513 GATGCAATTCATCATCTGTAAGCAAAATCTGCTGTAATTGAATCTACCT 562
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
200 roAlaGlyGlyGlyLysAspGluGlnGlnLeuValGluAlaPheGluGlu 216
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
563 CTTTAGCT.....CAAGATTCATGCTGAATTTTCAAT 597

```

```

217 MetThrArgAsnLeuPheSerLeuProIleAspValProPheSerGlyLeu 233
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
598 CTAAGCTCTAGGACGCTCTTTCACCTACTATCAATCTTCCAAAGCAACATA 647
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
233 uTyrArgGlyValLysAlaArg.....AsnLeuIleHisAla 246
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
648 TCATCCCGGATTTACAGCAAGAAATTAATTTGTAACCTATTACGAACAC 697
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
246 rGluGluGluAsnIleArgAlaLysIleArgArgLeuGlnIleThrGlu 262
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
698 TCATAGAACAG.....AGAGAGCTTCAAGAGAAATTCAA..... 732
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
263 ProAspGlyCysLysAspAlaLeuGlnLeuLeuIleGlnHisSerTr 279
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
733 .....CATGATATGCTTGGTTACCTGATGAATGAGAGACG 767
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
279 pGluArgGlyIuArgLeuAspMetGlnAlaLeuLysGlnSerSerThg 296
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
768 AACAGATTCAAATTAACAGATGATGATGATGATTAATTAATTAATTA 816
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
296 LuLeuLeuPheGlyIuHisGluThrThrAlaSerAlaIleThrSerLeu 312
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
817 ..ATTTGTACTCTGGATATGAACCTGTTCCACCACTTCTATGATGCT 864
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
313 IleThrTyrLeuGlyLeuTyrProHisValLeuGlnLysValArgGlu 329
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
865 GGAATATATCTTCATGATCATCCAAAGTTCTTGAAGAACTTAGAATAA 914
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
329 uIleLysSerLysGlyLeuLeuCysLysSerAsnGlnAspAsnLysLeu 346
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
915 A.....CACATGGCTATTAGAGAAAGAAAGAAACCTGAGATCCATG 958
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
346 sPMeGluThrLeuGlnGluGlnLysTyrIleGlyCysValIleLysGlu 362
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
959 ATTACACGATTTACAGGTCATAGCGGTTCAACAGACCTGATTTAGAG 1008
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
363 ThrLeuArgLeuAsnProValProGlyGlyPheArgValAlaLeuLys 379
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1009 ACCCTCAGGTTAGCAACATAGTAATGGGGTTTGAGAAAAACAACCTCA 1058
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
379 sThrPheGluLeuAsnGlyTyrGlnIleProLysGlyTyrPAsnValIle 396
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1059 AGATATGGAATAAATAGCGTATATCATTCCTTAAGAGATGAGAAATATAC 1108
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
396 yrSerIleCysAspThrHisAspValAlaAspIlePheThrAsnLysGlu 412
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1109 TATATACAAAGGAGTTGAATTAACGATCCAAAGACTTATCTGATCCATAT 1158
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
413 GluPheAsnProAspArgPheIleValProHisProGluAspAlaSerAr 429
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1159 TCGTTCATATCCATGGAGATGGATGATAGAGCCCTGGAACACCAAAAC.. 1206
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
429 gPheSerPheIleProPheGlyGlyLeuArgSerCysValGlyLysG 446
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1207 ...TCATTTTGGTATTGTGAGGGGTGCTAGACAAATGCTCTGGAAAG 1252
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
446 LuPheAlaLysIleLeuLeuLysIlePheThrValGluLeuAlaArgHis 462
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1253 AACTTGCTAGACAGAAATTTCCACATTTCTTCATTTACTTCTGTAACAAA 1302
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
463 CysAspTrpGlnLeuLeuAsnGlyProProThrMetLys 475
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1303 TACAGATGGGAAGAAATAGTGGAGATAAATCATGATGA 1341

```


Align seg 1/1 to: AAV09252 from: 1 to: 1725

```

1 MetGlyLeuProAlaLeuLeuAlaSerAlaLeuGlyThrPheValLeuPr 17
  |||
25 ATGGGGGCTCCGGCGCTGCCAGTGGCTGGCTGCACCTTCGTGCTGCC 74
  |||
17 OleuLeuLeuPheLeuAlaAlaLeuLysLeuTrpAspLeuTyrCysValS 34
  |||
75 GCTGCTGCTCTTCGGGGGGCTCAACCTCGGAGCTGACTGTGTGTA 124
  |||
34 erSerAlaGspArgSerCysAlaLeuProLeuProGlyThrMetGly 50
  |||
125 GCACCCGGGATCGCAGCTGCCCTCCCTGCCCGCCGGTACCATGGGC 174
  |||
51 PheProPhePheGlyLeuThrLeuGlnMetValLeuGlnArgArgLysPh 67
  |||
175 TTCCTATTCTTTGGGGAAACATGCAAGTGGTCTTCAGCGAGGAAAGTT 224
  |||
67 eLeuGlnMetLysArgArgLysTyrGlyPheIleTyrLysThrHisLeuP 84
  |||
225 TCTGCAGATGAAAGCGCAAGGAATACGGCTTATCATCAAGACGATCTGT 274
  |||
84 heGlyArgProThrValArgValMetGlyAlaAspAsnValArgArgIle 100
  |||
275 TTGGCGCGCCACGCGTGGGTGATGGCGCGGATTAATGTCGGCGCATC 324
  |||
101 LeuLeuGlyGlnHisArgLeuValSerValHisTrpProAlaSerValAr 117
  |||
325 TTGCTGGAGAGACACGCGTGGTGTGGTGCATGGCCCGCTGGTGGCG 374
  |||
117 gTrhIleLeuGlyValaGlyCysLeuSerAsnLeuHisAspSerHisL 134
  |||
375 CACCATCTGGGCGCTGGCTGCTCTCCAACTCGACATTCCTCGGCACA 424
  |||
134 ysgInaGlyLysValIleMetGlnAlaPheSerArgLysIleLeuGln 150
  |||
425 AGCAGCGAAAGAGTATTAATGACAGGCTTCACCGCGAGCATCTCCAG 474
  |||
151 CysTrpValLeuValIleAlaGlnGluValSerSerCysLeuGlnGlnTr 167
  |||
475 TGTGACTGCTCGTGCATCGCTGAGGAAGTCAAGCATGTCTGGAGCACTG 524
  |||
167 pLeuSerCysGlyLysValArgGlyLeuLeuValTyrProGluValLysArgL 184
  |||
525 GCTAAGCTGGCGGAGCGCGCTCTGCTACCCGAGGTGAAGCGCC 574
  |||
184 eueMetPheArgIleAlaMetArgIleLeuLeuGlyCysGlnProGlyPro 200
  |||
575 TCATGTTCCGATCGCATCGCATCGCTGCTGGGCTCGAGCCGGGTCCA 624
  |||
201 AlaGlyGlyLysLysAspGlnGlnGlnLeuValGlnIleAlaPheGlnGluMe 217
  |||
625 GCGGGCGGGGAGGAGCAACAACACTCGTGGAGCTTCGAGGAGAT 674
  |||
217 tThrArgAsnLeuPheSerLeuProIleAspValProPheSerGlyLeuT 234
  |||
675 GACCCGCAATCTTCTCTCTCCATTTGACGTGCCCTTTAGCGGCTGT 724
  |||
234 yTrArgGlyValLysAlaArgAsnLeuIleHisAlaArgIleGlnGlnAsn 250
  |||
725 ACCGGGGCGGAAAGCGCGCAACCTTATACAGCGCGCATCGAGGAAC 774
  |||
251 IleArgAlaLysIleArgArgLeuGlnAlaThrGlnProAspGlyLysC 267
  |||
775 ATTCGGCGCAAGATCCGCGCTTCAAGGCTACAGAGCCGATGGGGTTG 824
  |||
267 sLysAspAlaLeuGlnLeuLeuIleGlnHisSerTrpGlnArgGlyLysL 284
  |||
825 CAAGAGCGGCTCGCTCTCTGATTAAGCACTCGTGGGAAGGGAAGA 874
  |||
284 rGluAspMetGlnAlaLeuLysGlnSerSerThrGlnLeuPheGly 300
  |||

```

```

875 GGCTGATATGCGAGGCACTAAACAATCGTCAACAGAGCTCTCTTGGT 924
  |||
301 GlyHisGlnThrThrAlaSerAlaAlaThrSerLeuIleThrTyrLeuG 317
  |||
925 GGTCAATGAACATACAGCAGTGTCTCGACATCACTGATCACTTACCTAG 974
  |||
317 yLeuTrpProHisValLeuGlnLysValArgGlnGlnIleLysSerLysG 334
  |||
975 ACTTACCCACATGTTCTTCCAGAAAGTTTCAGAGAGATTAAGAAGCAAG 1024
  |||
334 lYleuLeuCysLysSerAsnGlnAspAsnLysLeuAspMetGlnThrLeu 350
  |||
1025 GCTTACTTTGCAAGAGCAATCAAGACAACAAGTTATGACATGGAAACTTTG 1074
  |||
351 GluGlnLeuLysTyrIleGlyCysValIleLysGlnThrLeuArgLeuAs 367
  |||
1075 GAACAGCTTAATATACATTGGGTGTCTCATTTAAGGAGACCTTCGATTAA 1124
  |||
367 nProProValProGlyGlyPheArgValAlaLeuLysThrPheGlnLeuA 384
  |||
1125 TCCTCCGTTCCAGGAGGCTTCCGGTGTCTGTGAGACTTTTGAGCTGA 1174
  |||
384 snGlyTyrGlnIleProLysGlyTrpAsnValIleTyrSerIleCysAsp 400
  |||
1175 ATGGATACCAAGATCCCAAGGCGTGAATGTTATTACAGTATGTGTGAC 1224
  |||
401 ThrHisAspValAlaAspIlePheThrAsnLysGlnGluPheAsnProAs 417
  |||
1225 ACCCAGATGTGGCAGATATCTTCAACAAGAGAGATTAATCCCGA 1274
  |||
417 PARPheIleValProHisProGluAspAlaSerArgPheSerPheIleP 434
  |||
1275 CCGCTTATATGCTCTCATCCAGAGATGCTTCCCGGTTCAGCTTCATTC 1324
  |||
434 ropheGlyGlyLysLeuArgSerCysValGlyLysGlnPheAlaLysIle 450
  |||
1325 CATTTGGAGAGGCGCTTCGAGCTGTGTAGGCAAAAGATTGCAAAATT 1374
  |||
451 LeuLeuLysIlePheThrValGlnLeuAlaArgHisCysAspTrpGlnLe 467
  |||
1375 CTCTTAAATATTATACAGTGAAGCTGGCTGACACTGTGATTTGGACGT 1424
  |||
467 uLeuAsnGlyProProThrMetLysTrpHisSerProThrValTyrProValA 484
  |||
1425 TCTAATAGGACCTCTCAATAGAGAGACAGCCCACTGTATACCTGTGG 1474
  |||
484 sPAsnLeuProAlaArgPheThrTyrPheGlnGlyAspIle 497
  |||
1475 ACAATCTCCCTGCAAGATTCACCTTACCTTCAGAGGAGATATTC 1515
  |||
seq_name: /SIDSI/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV12205
seq_documentation_block:
ID: AAV12205 standard; cDNA; 1725 BP.
XX
XX AAV12205;
XX
DT 22-JUN-1998 (first entry)
XX
DE Mouse retinoid metabolizing protein mP450RA1 cDNA.
XX
KW Retinoid metabolizing protein; P450RA1; retinoid oxidase;
KW retinoid acid; mouse; inhibitor; antisense; cancer;
KW actinic keratosis; oral leukoplakia; head tumour; neck tumour;
KW non-small cell lung carcinoma; basal cell carcinoma;
KW acute promyelocytic leukaemia; skin cancer; acne; psoriasis;
KW ichthyosis; therapy; diagnosis; screening; cytochrome P450; ss.
XX
OS Mus musculus.
XX
Key Location/Qualifiers
FT 25..1518
FT CDS /tag= a
FT

```

XX W09749815-A1.
 XX 31-DEC-1997.
 XX 23-JUN-1997; 97MO-CA00440.
 XX 01-OCT-1996; 96US-0724466.
 XX 21-JUN-1996; 96US-0667546.
 XX (TOOH) UNIV QUEBENS KINGSTON.
 PA Beckett BR, Jones G, Petkovich PM, White JA;
 F1 WPI; 1998-077178/07.
 DR P-PSDB; AAM44161.
 XX Retinoid metabolizing protein - useful to develop products to treat,
 PT e.g. cancer, acinic keratosis, oral leukoplakia, acne, psoriasis or
 PT Ichthyosis
 PS Claim 15; Page 62-64; 110pp; English.
 XX This cDNA clone codes for a novel mouse retinoid metabolizing
 CC protein (see AAM44160) designated mp450RAI. The encoded protein is a
 CC retinoid oxidase that has the ability to hydroxylate retinoic acid
 CC at the 4 position of the beta-ionone ring and is inducible in
 CC epithelial cells exposed to retinoic acid. The clone was isolated
 CC from a retinoic acid-treated P19 teratocarcinoma cDNA library
 CC using human P450RAI cDNA (see AAV12204) as probe. Zebrafish, human
 CC and mouse P450RAI cDNA clones (see AAV12203-05) are claimed. They
 CC can be used in the recombinant production of P450RAI proteins.
 CC Antisense nucleic acids can be used in a claimed method for
 CC inhibiting retinoic acid hydroxylation for the treatment of cancer,
 CC acinic keratosis, oral leukoplakia, secondary tumours of the head
 CC and/or neck, non-small cell lung carcinomas, basal cell carcinomas,
 CC acute promyelocytic leukaemia, skin cancer, and premalignancy
 CC associated with acinic keratosis, acne, psoriasis and/or
 CC Ichthyosis. Promoter sequences (see AAV12206-08) for the P450RAI
 CC genes are also claimed.
 XX Sequence 1725 BP; 408 A; 425 C; 475 G; 417 T; 0 other;
 SO
 alignment_scores: Length: 497
 Quality: 2612.00 Gaps: 0
 Ratio: 5.256 Percent Identity: 100.000
 Percent Similarity: 100.000 Percent Identity: 100.000
 alignment_block:
 US-09-668-482-32 x AAV12205 ..
 Align seg 1/1 to: AAV12205 from: 1 to: 1725
 1 MetGlyLeuProAlaLeuAlaSerAlaLeuCysThrPheValLeuP 17
 25 ATGGGGCTCCGGGCTGGCTGGCAAGTCCCTGCACTTCCTGCTGCC 74
 17 OLeuLeuPheLeuAlaAlaLeuLysLeuTyrPaspLeuTyrCysValS 34
 75 GCTGCTCTCTTCTCTGGCGGCTCAAGCTCTGGGACCTGTACTGTGTGA 124
 34 eSerArgAspArgSerCysAlaLeuProLeuProProGlyThrMetGly 50
 125 GCACGCCCGATCGACGCTGGCCCTCCCTTGGCCCCCGGTACCAGGGC 174
 51 PheProPhePheGlyGluThrLeuGlnMetValLeuGlnArgArgLysP 67
 175 TTCCCATCTTTTGGGGAACATTCACATGATGCTTCAGGGAGAGATT 224
 67 eLeuGlnMetLysArgArgLysTyrGlyPheIleTyrLysThrHisLeu 84
 225 TCTGCAGATGAAGCGCAGAAATACGGCTTCATCTACAGACGATCTGT 274

84 heGlyArgProThrValArgValMetGlyAlaAspAsnValArgArgIle 100
 275 TTGGGCGGCCACCGTCCGGGTGATGGGCGCGGATATATGCGCGCATC 324
 101 LeuLeuGlyGlnHisArgLeuValSerValHisTrpProAlaSerValAr 117
 325 TTGCTGGGAGAGACACCGTTGGTGTGCTGCACACTGCGCCCGTGGCG 374
 117 gThrIleLeuGlyAlaGlyCysLeuSerAsnLeuHisAspSerSerIst 134
 375 CACCATCTGGGCGCTGGCTGCCCTCCCAACCTGCACAGATTCTCCGACA 424
 134 YsgLinarGlyLysValIleMetGlnAlaPheSerArgGlnAlaLeuGln 150
 425 AGCAGCGAAGAGAGGTGATTATTCAGAGCCCTTCAGCCCGAGGACCTCAG 474
 151 CysTyrValLeuValIleAlaGluGluValSerSerCysLeuGlnItr 167
 475 TGCTACGTGCTCGTATCGCTGAGGAAGTCAGAGTTGTCTGGAGCAGTG 524
 167 pLeuSerCysGlyGluArgGlyLeuLeuValTyrProGluValLysArgL 184
 525 GCTAAGCTGCGGCGAGCGCGGCTCTGCTTACCCTCCAGGTGAAGCGCC 574
 184 eumePheArgIleAlaMetArgIleLeuLeuGlyCysGluProGlyPro 200
 575 TCATGTTCCGATGCGCATCGGCATCTCTGCTGGCTGGAGCCGGGTCCA 624
 201 AlaGlyGlyGlyGluAspGluGlnGlnLeuValGlnAlaPheGluJme 217
 625 GCGGGCGCGGGGAGAGACGACCAACAGCTCTGGAGGCTTCGAGAGAT 674
 217 tThrArgAsnLeuPheSerLeuProIleAspValProPheSerGlyLeu 234
 675 GACCGGCATCTCTCTCTCTCCATTGACGTGCGCTTGAAGCGCGTGT 724
 234 YrArgGlyValLysAlaArgAsnLeuIleHisAlaArgIleGluGluAs 250
 725 ACCGGGCGGTGAAGCGCGGGAACCTTATACACGCCGCGATGAGAGAAC 774
 251 IleArgAlaLysIleArgArgLeuGlnAlaThrGluProAspGlyGlyC 267
 775 ATTCCGCCCAAGATCCGCCGCTTCAGGCTACAGAGCGGATGGGGTTG 824
 267 sLysAspAlaLeuGlnLeuLeuIleGlnHisSerTrpGluArgGlyGlu 284
 825 CAAGAGCGCGCTGCAGCTCTGATTGACACCTCGTGGAGAGGGAGAGA 874
 284 rGLeuAspMetGlnAlaLeuLysGlnSerSerThrGluLeuPheGly 300
 875 GGCTGATATGCAAGCAGCAGCAAAACAATCGTCAACAGACTCTCTTGGT 924
 301 GlyHisGluThrThrAlaSerAlaAlaThrSerLeuIleThrTyrLeuG 317
 925 GGTCAATGAACATACAGCCAGTCTCGACATCAGTCACTTACTTGGG 974
 317 YLeuTyrProHisValLeuGlnLysValArgGluGlnIleLysSerLys 334
 975 ACTCTACCCACATGTCCTCAGAAAGTTTCGAGAAAGATAAAGACCAAG 1024
 334 YLeuLeuCysLysSerAsnGlnAspAsnLysLeuAspMetGluThrLeu 350
 1025 GCTTACTTTGCAAGAGCATCAAGACAACAAGTATGACATGGAACATTGG 1074
 351 GluGlnLeuLysTyrIleGlyCysValIleLysGluThrLeuArgLys 367
 1075 GAACACCTTAATAATACATTTGGGTGTGCATTAAGAGACCTCGATTGA 1124
 367 nProProValProGlyGlyPheArgValAlaLeuLysThrPheGluLeu 384
 1125 TCTCTCGGTTCCAGAGGGTTTCGGGTCTCTGAAGACTTTTGAAGCTGA 1174

```

384  snGlyTyrGlnIleProLysGlyTyrPasnValIleTyrSerIleCysAsp 400
      |||
1175  ATGGATACCAAGATCCCAAGGGCTGGAATGTTATTACATCTCTGTGAC 1224
      |||
401  ThrHisAspValAlaIleAspIlePheThrAsnLysGluIlePheAsnProAs 417
      |||
1225  ACCCAGAGATGGCAGATATCTTCACTAACACAGAGAAATTTATCCCGA 1274
      |||
417  ParPheIleValProHisProGluAspAlaSerArgPheSerPheIleP 434
      |||
1275  CGGCTTATAGTGCCTCATCCAGAGATGCTTCCCGTTCCAGCTTCATTC 1324
      |||
434  ropHegIlyGlyIleuArgSerCysValGlyLysGluPheAlaLysIle 450
      |||
1325  CATTTGGAGAGAGCCCTTCGAGCTGTGTAGCAAGAGTTTGCAAAATTT 1374
      |||
451  IleuLysIlePhePheValGluIleuAlaArgHisCysAspTyrGlnIle 467
      |||
1375  CTTCCTAAGATATTACATGAGTGGCTGCTAGGCACTGTGATGGCAGCT 1424
      |||
467  uLeuAsnGlyProProThrMetLysThrSerProThrValTyrProValA 484
      |||
1425  TCTAATATGACCTCTACATGACAGACAGCCCACTGTGTACCTGTGG 1474
      |||
484  SPAsnLeuProAlaArgPheThrTyrPheGlnGlyAspIle 497
      |||
1475  ACAATCTCCCTGCAAGATTCACCTACTTCCAGGAGATATC 1515
      |||

```

seq_name: /SIDSL/gcgdata/geneseq/geneseq/NAL198.DAT:AAV09247

seq_documentation_block:

ID AAV09247 standard; cDNA; 1494 BP.

AAV09247;

07-JUL-1998 (first entry)

Human cytochrome P450RAI cDNA sequence.

Retinoid regulated gene; cytochrome P450 gene; enzyme;

oxidative metabolism; P450RAI; retinoic acid; RA; promoter; ss.

Homo sapiens.

Key Location/Qualifiers

CDS 1..1494

/*tag- a

/product- "Cytochrome P450RAI"

MO9749832-A2.

31-DEC-1997.

23-JUN-1997; 97WO-CA00488.

01-OCT-1996; 96US-0724466.

21-JUN-1996; 96US-0667546.

(TOOH) UNIT QUEENS KINGSTON.

Petkovich PM;

WPI; 1998-077193/07.

P-PSDB; AAW37734.

Identifying DNA encoding inducible or suppressible cytochrome P450 -
by screening for drugs which reduce the catabolism of retinoic acid,
useful in cancer chemotherapy and the treatment of acne and
psoriasis

Example 4; Pages 56-58; 113pp; English.

This nucleotide sequence encodes the human cytochrome P450RAI. Its

CC expression is dependent on the presence of retinoic acid (RA). The
CC retinoid-regulated genes such as the inducible cytochrome P450RAI
CC gene specifically metabolizes a derivative of the RA. The cytochrome
CC P450 gene in general produces enzymes involved in the oxidative
CC metabolism of endogenous and exogenous compounds. The cytochrome
CC P450 nucleotide sequence can be used to induce or suppress the
CC expression of its protein. P450RAI is highly induced by RA in cell
CC lines and tissues. This allows for development of a drug screen
CC using promoters and nucleotide sequences to identify drugs which are
CC useful for reducing the catabolism of RA.
XX

Sequence 1494 BP; 332 A; 414 C; 430 G; 318 T; 0 other;

alignment_scores:

Quality: 2437.00	Length: 497
Ratio: 5.035	Gaps: 0
Percent Similarity: 97.384	Percent Identity: 93.360

alignment_block:

US-09-668-482-32 x AAV09247 ..

Align seg 1/1 to: AAV09247 from: 1 to: 1494

```

1 MetGlyLeuProAlaIleuAlaSerAlaLeuCysThrPheValLeuPr 17
  |||
1 ATGGGCTCTCCCGGCTGCTGGCCATGCGCTGTGACCTTGCTGTGGC 50
  |||
17 oLeuLeuPheLeuAlaIleuLysLeuTyrPaspLeuTyrCysValS 34
  |||
51 GCTGCTGCTCTTCTCGTGGCTGCACTGAGCTGGAGCTGTACTGTGTA 100
  |||
34 eSerArgAspArgSerCysAlaLeuProLeuProGlyThrMetGly 50
  |||
101 GCGGCGCGCAGCCAGTTGTGCTCCATTCCTCCCGGCACTATGGGC 150
  |||
51 PheProPheGlyIleuThrLeuGlnMetValLeuGlnArgArgLysPh 67
  |||
151 TTCCTCTTCTTGGGAAACCTTGCAATGATGTAAGGAGGAGATT 200
  |||
67 eLeuGlnMetLysArgArgLysTyrGlyPheIleTyrIleThrHisLeuP 84
  |||
201 CCTGCAGATGAAGCGCAGGAATACGGCTTCATCTACAAGACGCAATGT 250
  |||
84 heGlyArgProThrValArgValMetGlyAlaAspAsnValArgArgIle 100
  |||
251 TCGGGCGGCCACCGCTACGGGTGATGGCGGACAAATGTGGCGGCATC 300
  |||
101 LeuLeuGlyGlnHisArgLeuValSerValHisTrpProAlaSerValAr 117
  |||
301 TTGCTCGGAGAGACGCGGCTGGTGTGCGTCCACTGGCCAGCGTGGTGG 350
  |||
117 gThrIleLeuGlyAlaGlyCysLeuSerAsnLeuHisAspSerSerHisL 134
  |||
351 CACCATTCCTGGGATCTGGCTGCTCTTACCTGCACAGCTCTTCGCACA 400
  |||
134 ySglnArgLysLysValIleMetGlnAlaPheSerArgGlnAlaLeuGln 150
  |||
401 AGCAGCGCAAGAGGTGATATGCGGCTTCAGCGGAGGACCTCGCA 450
  |||
151 CysTyrValLeuValIleAlaGlnGlnValSerSerCysLeuGlnIleTr 167
  |||
451 TGCTACGTCGCGGTGATACCGAGGAATGGGACAGCTGGAGCACTG 500
  |||
167 PleuSerCysGlyLeuArgGlyLeuLeuValTyrProGluValLysArgL 184
  |||
501 GCTGAGTCGCGGAGCGCGCTCTGCTTACCCAGGTGAAGCGCC 550
  |||
184 eumetPheArgIleAlaMetArgIleLeuLeuGlyCysGluProGlyPro 200
  |||
551 TCATGTTCCGAATCGCATCGCATCTACATGCGGTGGAAGCCCACTG 600
  |||
201 AlaGlyGlyGlyGluAspGlnGlnIleuValGlnAlaPheGlnGlu 217

```

```

||||| |||.....|
601 GCGGCGCAGCGGAGCTCCGACGACAGCTGTGTGGAGGCTTCGAGGAAT 650
217 tThrArgAsnLeuPheSerLeuProIleAspValProPheSerGlyLeu 234
||||| |||.....|
651 GACCGGCAATCTCTTCTGCTGCCCATGAGCTGCCCTTCAGCGGCTGT 700
234 yTAAGGlyValLysAlaArgAsnLeuIleHisAlaArgIleGluGluAsn 250
||||| |||.....|
701 ACCGGGCGATGAGGCGGGAACCTCATTTCAAGCGCGCATCGACAGAAC 750
251 IleArgAlaLysIleArgArgLeuGlnAlaThrGluProAspGlyGlyCy 267
||||| |||.....|
751 ATTCGCGCCAGATCTCGGCGCTGCGGCAATCCAGCGGCGCCAGGCGCTG 800
267 sLysAspAlaLeuGlnLeuIleGluHisSerTrpGluArgGlyLys 284
||||| |||.....|
801 CAAGAAGCGCGTGCAGCTGTTCAGATCGACACTCGTGGAGAGGAGAGC 850
284 rGluAspMetGlnAlaLeuLysGlnSerSerThrGluLeuPheGly 300
||||| |||.....|
851 GCGTCGACATGACGACACTAAAGCAATCTTCACCGCACTCCTTTTGA 900
301 GlyHisGluThrThrAlaSerAlaIleThrSerLeuIleThrTyrLeuG 317
||||| |||.....|
901 GGACAGGAACCGCGCCAGTGCACGCCACATCTCGATCACTTACTCTGG 950
317 yLeuTyrProHisValLeuGlnLysValArgGluGluIleLysSerLys 334
||||| |||.....|
951 GCTTACCCACATGTTCTCCAGAAAGTCGAGAAAGACTGAAGAGTAAG 1000
334 yLeuLeuLysLysSerAsnGlnAspAsnLysLeuAspMetGluThrLeu 350
||||| |||.....|
1001 GTTCTCTTTGCAAGAGCAATCAAGACAACAAAGTTGACATGGAAATTTTG 1050
351 GluGlnLeuLysTyrIleGlyCysValIleLysGluThrLeuArgLeuAs 367
||||| |||.....|
1051 GAAACAACCTTAATACATCGGCTGTGTATTAAAGACCCCTTCGACTGA 1100
367 nProProValProGlyGlyPheArgValAlaLeuLysThrPheGluLeu 384
||||| |||.....|
1101 TCCCCAGATTCAGAGGAGGTTTCGGGTTCTCTGAAGACTTTTGAATTA 1150
384 sngGlyTyrGlnIleProLysGlyTTPAsnValIleTyrSerIleCysAsp 400
||||| |||.....|
1151 ATGGATACCAAGATCCCAAGGCGTGAAGTATCTACAGTATCTGTGAT 1200
401 ThrHisAspValAlaAspIlePheThrAsnLysGluGluPheAsnProAs 417
||||| |||.....|
1201 ACTCATGATGTGGCAGAGATCTTCACCAACAAGAAATTTAAATCTGA 1250
417 pArgPheIleValProHisProGluAspAlaSerArgPheSerPheIleP 434
||||| |||.....|
1251 CCGATTCAGTGGCTCTCACCAAGAGATCATTCACAGTTCACTTCATATC 1300
434 rPheGlyGlyGlyLeuArgSerCysValGlyLysLubPheAlaLysIle 450
||||| |||.....|
1301 CATTGTGAGAGGCTTAGAGCTGTGTAGGCAAAAGAAATTTGCAAAAAAT 1350
451 LeuLeuLysIlePheThrValGluLeuAlaArgHisCysAspTrpGlnLe 467
||||| |||.....|
1351 CTTCTCAAAATATTTACATGGAGCTGGCCAGGCAATGTGACGCGAGCT 1400
467 uLeuAsnGlyProProThrMetLysThrSerProThrValTyrProValA 484
||||| |||.....|
1401 TCTAAATGAGACTCTCAATGAAAAACAGTCCACCGTATACCTCTGTGG 1450
484 sPAsnLeuProAlaArgPheThrTyrPheGlnGlyAspIle 497
||||| |||.....|
1451 ACAATCTCCCTGCAGAGATTACCCATTTCCATGGGGAATC 1491
seq_name: /SIDSI/gcgdata/geneseq/geneseqn/NAI98.DAT:AAV12204

```

```

seq_documentation_block:
ID AAV12204 standard; cDNA; 1494 BP.
XX
AC AAV12204;
XX
DT 22-JUN-1998 (first entry)
XX
DE Human retinoid metabolising protein hp450RAI cDNA.
XX
KW Retinoid metabolising protein; P450RAI; retinoid oxidase;
KW retinoid acid; human; inhibitor; antisense; cancer;
KW actinic keratosis; oral leukoplakia; head tumour; neck tumour;
KW non-small cell lung carcinoma; basal cell carcinoma;
KW acute promyelocytic leukaemia; skin cancer; acne; psoriasis;
KW ichthyosis; therapy; diagnosis; screening; cytochrome P450; ss.
XX
OS Homo sapiens.
XX
PN MO9749815-A1.
XX
PD 31-DEC-1997.
XX
PE 23-JUN-1997; 97MO-CA00440.
XX
PR 01-OCT-1996; 96US-0724466.
PR 21-JUN-1996; 96US-0667546.
XX
PA (TOOH ) UNIV QUEENS KINGSTON.
XX
PI Beckett BR, Jones G, Petkovich PM, White JA;
XX
DR WPI: 1998-077178/07.
XX
DR P-PSDB; AAM44160.
XX
PT Retinoid metabolising protein - useful to develop products to treat,
PT e.g. cancer, actinic keratosis, oral leukoplakia, acne, psoriasis or
PT ichthyosis
XX
PS Claim 15, Page 55-57; 110pp; English.
XX
CC This cDNA clone codes for a novel human retinoid metabolising
CC protein (see AAM44160) designated hp450RAI. The encoded protein is a
CC retinoid oxidase that has the ability to hydroxylate retinoid acid
CC at the 4 position of the beta-ionone ring and is inducible in
CC epithelial cells exposed to retinoid acid. To isolate the clone,
CC zebrafish P450RAI was used to search an expressed sequence tag
CC database, and an isolated clone was used to screen a cDNA library
CC generated from an NT2 cell line treated with retinoid acid. The
CC hp450RAI gene has been localised to 10q23-24. Zebrafish, human
CC and mouse P450RAI cDNA clones (see AAV12203-05) are claimed. They
CC can be used in the recombinant production of P450RAI. Antisense
CC nucleic acids can be used in a claimed method for inhibiting
CC retinoid acid hydroxylation for the treatment of cancer, actinic
CC keratosis, oral leukoplakia, secondary tumours of the head and/or
CC neck, non-small cell lung carcinomas, basal cell carcinomas, acute
CC promyelocytic leukaemia, skin cancer, and premalignancy associated
CC with actinic keratosis, acne, psoriasis and/or ichthyosis.
CC Promoter sequences (see AAV12206-08) for the P450RAI genes are also
CC claimed.
XX
SQ Sequence 1494 BP; 332 A; 414 C; 430 G; 318 T; 0 other;
XX
alignment_scores:
Quality: 2437.00 Length: 497
Ratio: 5.035 Gaps: 0
Percent Similarity: 97.384 Percent Identity: 93.360
alignment_block:
US-09-668-482-32 x AAV12204 ..
Align seg 1/1 to: AAV12204 from: 1 to: 1494

```

```

1 MetGlyLeuProAlaLeuAlaSerAlaLeuCysThrPheValLeuPr 17
  |||
1 ATGGGGCTCCCGGCTGCTGGCGAGTGGCTGTGACCTTCGTCGTGCC 50
17 OLeuLeuPheLeuAlaAlaLeuLysLeuThrPheLeuTyCysValS 34
  |||
51 GCTGCTGCTCTTCGTGGCTGGCATCAAGCTTGGGACCTGTACTGTGTA 100
34 erSerArgAspArgSerCysAlaLeuProLeuProProGlyThrMetGly 50
  |||
101 GGGGCGGAGCCGCGAGTTGGCTCCCATTTGCCCGCGGACTATGGGC 150
51 PheProPhePheGlyLutThrLeuGlnMetValLeuGlnArgArgLysPhe 67
  |||
151 TTCCCTCTTCTTGGGAAACCTTGACAGATGGTACGACCGGAGGACGTT 200
67 eLeuGlnMetLysArgArgLysTyGlyPheIleTyThrLysThrHisLeuP 84
  |||
201 CCTGCAGATGAAGCGCAGGAATACGGCTTCATCTACAAAGACGCAATCTGT 250
84 heGlyArgProThrValArgValMetGlyAlaAspAsnValArgArgTyle 100
  |||
251 TCGGGCGGCGCCACCTACGGGTGATGGGCGCGACAAATGTCCGGCGCATC 300
101 LeuLeuGlyGlnHisArgLeuValSerValHisTrpProAlaSerValAr 117
  |||
301 TTGCTCGGAGACGACCGGCTGGTGTCCACTGCGCCAGCGCTGGTGGG 350
117 gThrIleLeuGlnGlyAlaGlyCysLeuSerAsnLeuHisAspSerHisL 134
  |||
351 CACCATTTGGGATCTGGCTGCTCTTACCTGCACGACTCTCCGACAA 400
134 ySglnArgLysLysValIleMetGlnAlaPheSerArgLysAlaLeuGln 150
  |||
401 AGCAGCGCAAGAGTGATATACGGGCTTCAGCGGAGGACACTCCAA 450
151 CysThrValLeuValIleAlaGlnLysValSerSerCysLeuGlnGlyTr 167
  |||
451 TGTACGCGCGCGGTGATCCAGGAAAGTGGGACCGCTGGAGCACTGTG 500
167 pLeuSerCysGlyLysArgLysLeuValLeuValTrpProGlnValLysArgL 184
  |||
501 GCTAGCTGGCGGCGGCGGCTCTGCTACCCCGAGGTGAAGCGCC 550
184 euMetPheArgIleAlaMetArgIleLeuLeuGlyCysGluProGlyPro 200
  |||
551 TCAGTTTCCGATCGCATCGCATTCCTACTGGGCTGGCAACCCCAACTG 600
201 AlaGlyLysGlyLysAspGlnGlnLeuValGlnAlaPheGlnGlnLys 217
  |||
601 GCGGGCGAGCGGCGACCTCGAGCAGCAGCTTGTGAGAGCCCTTCAGGAAT 650
217 tThrArgAsnLeuPheSerLeuProIleAspValIleProPheSerGlyLeuT 234
  |||
651 GACCGCAATCTCTTCGGTGGCCATCGACGTCCCTTCAGCGGCGCTGT 700
234 yArgArgValLysAlaArgAsnLeuIleHisAlaArgIleGlnGlnAsn 250
  |||
701 ACCGGGGCATGAAGCGCGGCAACTCATTCACGGCGCATCGACAGAAC 750
251 IleArgAlaLysIleArgArgLeuGlnAlaThrGlnProAspGlyGlyCy 267
  |||
751 ATTGCGGCCCAAGATCTCGGGCTCGGGGCAATCCGAGCGCGGCGCCTG 800
267 sLysAspAlaLeuGlnLeuLeuIleGlnHisSerTrpGlnArgGlyLysAla 284
  |||
801 CAAGAAGCGCGCTGAGCTGTGATCGAGCACTCGTGGAGAGGAGGAGAC 850
284 rGlyLeuSerMetGlnAlaLeuLysGlnSerSerThrGlnLeuLeuPheGly 300
  |||
851 GGCTGGACATGCAAGGCACTAAAGCAATCTCAACCGCAACTCTCTTGA 900
301 GlyHisGlnThrThrAlaSerAlaAlaThrSerLeuLeuIleThrTyLeuGln 317

```

```

  |||
901 GGACAGAAACCAACGCGCAGTGCAGCCACATCTGTGATCACTTACCTGGG 950
317 yLeuTrpProHisValLeuGlnLysValArgGlnGlnIleLysSerLysG 334
  |||
951 GCTTACCCACATGTTCTTCAGAAAGTGCAGAAAGCTGAAAGAGTAAAG 1000
334 lYLeuLeuCysLysSerAsnGlnAspAsnLysLeuAspMetGlnThrLeu 350
  |||
1001 GTTTACTTTGCAAGAGCAATCAAGACAACAGATTGGACATGGAATTTTG 1050
351 GlnGlnLeuLysTyrlleGlyCysValIleLysGlnThrLeuArgLeuAs 367
  |||
1051 GAACAACTTAATACATCGGGTGTATTAAAGAGACCTTCGACTGAA 1100
367 nProProValProGlyGlyPheArgValAlaLeuLysThrPheGlnLeuA 384
  |||
1101 TCCCCAGTTCAGAGGAGGTTTCGGGTTGCTCTGAAGACTTTTGAATTA 1150
384 sngLyTyrglnIleProLysGlyTrpAsnValIleTySerIleCysasp 400
  |||
1151 ATGGATACCAAGATCCCAAGGCGTGAATGTATCTACAGTATCTGTGAT 1200
401 ThrHisAspValAlaAspIlePheThrAsnLysGlnGlnPheAsnProAs 417
  |||
1201 ACTCATGATGTGGCAGAGATCTTCACCAACAGAAAGATTTAATCTTGA 1250
417 pArgPheIleValProHisProGlnAspAlaSerArgPheSerPheIleP 434
  |||
1251 CCGATTCAAGTGCCTCCACCCAGAGATGCATCCAGTTCACTTCATTC 1300
434 ropPheGlyGlyLeuArgSerCysValGlyLysGlnPheAlaLysTyle 450
  |||
1301 CATTGGAGAGAGCCTTAGAGCTGTGTAGGCAAAAGATTTGCAAAATTT 1350
1351 CTTTCAAAATATTATACAGTGGAGCTGCCAGGCACTTGTGACTGCGACT 1400
451 LeuLeuLysIlePheThrValGlnLeuAlaArgHisCysAspTrpGlnLe 467
  |||
467 uLeuAsnGlyProProThrMetLysThrSerProThrValIleProValA 484
  |||
1401 TCTAAATGACCTCCTCAATGAATAAACCAGTCCACCGCTGTATCTGTGG 1450
484 sPAsnLeuProAlaArgPheThrTyrllePheGlnGlyAspIle 497
  |||
1451 ACAATCTCCCGCAAGATTCAACCATTTTCATGGGGAATC 1491
seq_name: /SIDSL/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA60752
seq_documentation_block:
ID AAA60752 standard; cdNA; 2124 BP.
XX
AC AAA60752:
XX
DT 31-OCT-2000 (first entry)
XX
DE Human PSEC64 encoding nucleotide sequence SEQ ID NO:1.
XX
KW Human; PSEC64; neuron growth; nerve disease; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 46..696
FT FT /tag= a "PSEC64"
FT FT /product= "PSEC64"
FT FT /note= "protein related to neuron growth"
XX
PD 06-JUN-2000.
XX
PF 19-NOV-1998; 98BP-0329989.

```


XX 19-NOV-1998; 98bp-0329989.
XX
XX (HERI-) HERIKKUSU KENKYUSHO KK.
PA
XX WPI: 2000-468126/41.
DR P-PSDB: AAB12489.
XX
XX A protein related to the growth of neuron and a gene encoding said
PT protein -
XX
PS Claim 2: Page 8-10; 13pp: Japanese.
XX
XX The present sequence encodes a human protein, designated PSEC64, which
CC is related to neuron growth. The PSEC64 protein and its gene can be used
CC for the development of a preventive agent for use in the treatment of
CC diseases in which nerves are involved.
XX
SQ Sequence 2124 BP; 509 A; 510 C; 547 G; 558 T; 0 other;

alignment_scores:
Quality: 2425.00 Length: 497
Ratio: 4.990 Gaps: 0
Percent Similarity: 97.787 Percent Identity: 93.763

alignment_block:
US-09-668-482-32 x AAA60752 ..

Align seg 1/1 to: AAA60752 from: 1 to: 2124

1 MetGlyLeuProAlaLeuAlaSerAlaLeuCystrPheValLeuP 17
46 ATGGGGCTCCGGCGCTGCTGCCAGTGGCTCTGCACCTTCGTGCTCC 95
17 OeuLeuPheLeuAlaAlaLeuLysLeuTrpAspLeuTyrCysValS 34
96 GCTGCTGCTCTTCCTGGCTGGATCAACCTCTGGACCTGTACTGCGTGA 145
34 eSerArgAspArgSerCysAlaLeuProLeuProGlyThyMetGly 50
146 GCGGGCGGACCGCAGTTGTGCTCCCATTCGCCCGGGGACTATGAGGC 195
51 PheProPhePheGlyGluThrLeuGlnMetValLeuGlnArgArgLysPhe 67
196 TTCCCTTCCTTTGGGAAACCTTGCAATGTGACTGCAGCGAGGAAAGTT 245
67 eLeuGlnMetLysArgArgLysTyrGlyPheIleTyrLysThrHisLeuP 84
246 CCTGCAGATGAAGCGCAGGAATACGGCTTCATCTACAAAGCATCTGT 295
84 heGlyArgProThrValArgValMetGlyAlaAspAsnValArgArgLle 100
296 TCGGGCGGCGCCACGCTACGGGTGATGGCGGACAAATGTGCGGCGCATC 345
101 LeuLeuGlyGlnHisArgLeuValSerValHisTrpProAlaSerValArg 117
346 TTGCTCGGAGACCGCGCTGGTGTGGTCCACTGGCCAGCGCTCGGTGCG 395
117 gThrIleLeuGlyAlaGlyCysLeuSerAsnLeuHisAspSerSerHisL 134
396 CACCAATTCGGGATCTGGTGTGCTCTTAACCTGCACGACTCTCGACAA 445
134 ysgLlnArgLysLysValIleMetGlnAlaPheSerArgLlnAlaLeuGln 150
446 ACACAGCGCAAGAGGATGATTCGCGGCTTCAGCCGCGAGGCACTCGAA 495
151 CysTyrValLeuValIleAlaGlnGluValSerSerCysLeuGlnGluTr 167
496 TGCTAGTCCGCTGATCACGAGAGAGTGGGACAGCTTGAGACAGTG 545
167 pleuSerCysGlyGluArgLysLeuLeuValTyrProGlnValLysArgL 184

546 GCTGACCTGGCGGAGCGCGGCTCTGCTACCCCGAGGTGAAGCGCC 595
184 euePheArgLleAlaMetArgLleLeuLeuGlyCysGlnProGlyPro 200
596 TCATGTTCCGAATCCGCATCGCATCTACTGGGCTGCGCAACCCAACTG 645
201 AlaGlyGlyGlnAspGlnGlnGlnValAlaGlnAlaPheGlnGluLeu 217
646 GCGGCGACGGGAGCTCCGAGCGACGCTTGTGGAGGCTTGAGAGA-AT 694
217 tThrArgAsnLeuPheSerLeuProIleAspValProPheSerGlyLeuT 234
695 GACCGCAATCTCTCTCGCTGGCCATCGACGTCCTTCAGGGGCTGT 744
234 YrArgLysValLysAlaArgAsnLeuIleHisAlaArgLleGlnGlnAsn 250
745 ACCGGGCAATGAAGGCGCGAACCTTACACCGCGCATTCGAGCGAAGC 794
251 IleArgAlaLysIleArgArgLysGlnAlaThrGlnProAspGlyGlyCy 267
795 ATTCCGCGCCAAAGATCTGGGCTGGCGGCATCCGAGCGCGCAGGGCTG 844
267 sLysAspAlaLeuGlnLeuLeuIleGlnHisSerTrpGlnArgLysGlnA 284
845 CAAGAAGCGGCTGACGCTGTGATCGACACTCGTGGAGAGGGAGAGAC 894
284 rGleuAspMetGlnAlaLeuLysGlnSerSerThrGlnLeuLeuPheGly 300
895 GCGTGGACATGACAGGCACTAAACCAATCTTCAACCGAACTCTTTTGA 944
301 GlyHisGlnThrThrAlaSerAlaAlaThrSerLeuIleThrTyrLeuG 317
945 GACACAGAAACCAAGCGCATGCGACGACATCTGTATCACTTACCTGG 994
317 yLeuTyrProHisValLeuGlnLysValArgLysGlnLleLysSerLysG 334
995 GCCTTACCCACATGTTCTCCAGAAAGTGCAGAAAGCTGAAAGTAAAG 1044
334 LysLeuLeuCysLysSerAsnGlnAspAsnLysLeuAspMetGluThrLeu 350
1045 GTTACTCTTGCAAGGCAATCAACACAAAGTTGGACATGGAAATTTTG 1094
351 GlnGlnLeuLysTyrIleGlyCysValIleLysGlnThrLeuArgLeuAs 367
1095 GAACAACCTTAATATCATCGGGTGTATTAAAGAGACCTTCGACAGTAA 1144
367 nProProValProGlyLysPheArgValAlaLeuLysThrPheGlnLeuA 384
1145 TCCCCAGTTCCAGAGAGGTTTCGGGTGCTGTGAAGACTTTTGAATTA 1194
384 snGlyTyrGlnIleProLysGlyTrpAsnValIleTyrSerIleCysAsp 400
1195 ATGGATACCCAGATTCACAAAGGCTGGAAATGTATTCTACATCTGTGAT 1244
401 ThrHisAspValAlaAspIlePheThrAsnLysGlnGluPheAsnProAs 417
1245 ACTCATGATGTGGCAGAGATCTTCACCAACAGAGAAATTAATCTGA 1294
417 pArgPheIleValProHisProGlnAspAlaSerArgPheSerPheIleP 434
1295 CCGATTCTATGCTGCCATCCACGAGAGATGCATCCAGGTTTCAGCTTCAT 1344
434 ropheGlyGlyLeuArgSerCysValGlyLysGlnPheAlaLysIle 450
1345 CATTTGGAGAGGCTTATAGAGCTGTGTAGCAAAAGATTTGCAAAATTT 1394
451 LeuLeuLysIlePheThrValGlnLeuAlaArgHisCysAspTrpGlnLe 467
1395 CTCTCAAAATATTACAGTGCAGCTGGCCAGCATTTGTGACTGCGAGCT 1444
467 uLeuAsnGlyProProThrMetLysThrSerProThrValTyrProValA 484
1445 TCTAAATGACCTCTCAATGAAGAAACCAAGTCCACCGGTATCTGTGG 1494

484 SPASnLeuProAlaArgPheThrTyrPheGlnGlyAspIle 497
|||||
1495 ACAATCTCCCTCGAAGATTCACCCATTCTCAGTCGGGAAATC 1535
seq_name: /SIDS1/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV12216
seq_documentation_block:
ID AAV12216 standard; DNA: 4164 BP.
XX
AC AAV12216;
XX
DT 22-JUN-1998 (first entry)
XX
DE Mouse retinoid metabolizing protein mp450RAI genomic DNA.
XX
KW Retinoid metabolizing protein; P450RAI; retinoid oxidase;
KW retinoic acid; mouse; inhibitor; antisense; cancer;
KW actinic keratosis; oral leukoplakia; head tumour; neck tumour;
KW non-small cell lung carcinoma; basal cell carcinoma;
KW acute promyelocytic leukaemia; skin cancer; acne; psoriasis;
KW ichthyosis; therapy; diagnosis; screening; cytochrome P450; ds.
XX
OS Mus musculus.
XX
PN M09749815-A1.
XX
PD 31-DEC-1997.
XX
PF 23-JUN-1997; 97MO-CA00440.
XX
PR 01-OCT-1996; 96US-0724466.
PR 21-JUN-1996; 96US-0667546.
XX
PA (TTOOH) UNIV QUEBENS KINGSTON.
XX
PI Beckett BR, Jones G, Petkovich PM, White JA;
XX
DR WPI; 1998-07178/07.
XX
PT Retinoid metabolizing protein - useful to develop products to treat,
PT e.g. cancer, actinic keratosis, oral leukoplakia, acne, psoriasis or
PT ichthyosis
XX
PS Disclosure; Page 69-70; 110pp; English.
XX
CC This sequence comprises a genomic clone of a novel mouse retinoid
CC metabolizing protein designated mp450RAI. A cDNA clone (see
CC AAV12205) that includes a coding sequence for the full-length
CC mp450RAI protein (see AAW44161) is also provided. The encoded
CC protein is a retinoid oxidase that has the ability to hydroxylate
CC retinoic acid at the 4 position of the beta-ionone ring and is
CC inducible in epithelial cells exposed to retinoic acid. Zebrafish,
CC human and mouse P450RAI cDNA clones (see AAV12203-05) are claimed.
CC They can be used in the recombinant production of P450RAI.
CC Antisense sequences can be used in a claimed method for inhibiting
CC retinoic acid hydroxylation for the treatment of cancer, actinic
CC keratosis, oral leukoplakia, secondary tumours of the head and/or
CC neck, non-small cell lung carcinomas, basal cell carcinomas, acute
CC promyelocytic leukaemia, skin cancer, and premalignancy associated
CC with actinic keratosis, acne, psoriasis and/or ichthyosis. Promoter
CC sequences (see AAV12206-08) for the P450RAI genes are also claimed.
XX
SQ Sequence 4164 BP; 997 A; 975 C; 1071 G; 1121 T; 0 other;

alignment_scores:
Quality: 1853.00 Length: 1046
Ratio: 4.240 Gaps: 10
Percent Similarity: 41.778 Percent Identity: 41.396

alignment_block:
US-09-668-482-32 x AAV12216 ..

Align seg 1/1 to: AAV12216 from: 1 to: 4164
40 CysAlaLeuProLeuProProGlyThrMetGlyPheProPhePheGly 56
|||||
73 TGGGCTCTCTCTCCCTCACCCTCCCTCT..... 99
56 uthrLeuGlnMetValLeuGlnArgArgLysPheLeuGlnMetLysArg 73
|||||
100CAGCGAGAGAAATTCTGCAAGATGAACGCA 130
73 rplysTyrGlyPheIleTyrLysThrHisLeuPheGlyArgProThrVal 89
|||||
131 GGAATACGGCTTCACTTACAAAGCCATCTGTTGGCGGCCACAGG 180
90 ArgValMetGlyAlaAspAsnValArgArgIleLeuLeuGlyGluHisAr 106
|||||
181 CGGGTATGCGCGCGGATATGCGCGCCATCTTCTGGAGAGCAGCG 230
106 gLeuValSerValHisTrpProAlaSerValArgThrIleLeuGlyAlaG 123
|||||
231 GTTGTGTGTGTGCTGCTGCGCGCGCTGCGCACCACATCTGGGGCTG 280
123 lCysLeuSerAsnLeuHisAspSerSerHisLysGlnArgLys..Lys.. 138
281 GCTGCTCTCCAACTGACGATCTCTCCACAGCAGCAAGAAAGGTG 330
138 138
331 AGGGTAGCTGGCAACTCTTGGCTGCGAGAGACCTCATCTAGTGGCT 380
138 138
381 TGGTTCAGCAAAATGAATGCGGGGCGAGGCTAGTCTATGTGTGGG 430
138 138
431 GACCAAGACCTCTCTATCTGATCCACTTATGCTTTCTGCTAGCAG 480
138 138
481 TGGGTAGTCTGCGGGGAGCTGAATTTCTGAAAGGTACTCGGAAGG 530
138 138
531 CGAAGGGGGGGGCTGAGGGAAGTACAGGATTGTACACTCTGCTC 580
139ValIleMetGlnAlaPheSerArgGluAlaLeuG 150
581 CTGGGGGTGCTCAGGTGATTATGACAGGCTTCACCGCGAGGCACTCA 630
150 nCysTyrValLeuValIleAlaGluGluValSerSerCysLeuGluGlnT 167
631 GTGCTACGTGCCCGGATGCTGCTGAGGAAAGTACAGTGTCTGGAGCAGT 680
167 rPLeuSerCysGlyLysArgGlyLeuLeuValTyrProGluValLysArg 183
681 GGGTAAAGCTGCGGCGAGCGGCTCTGCTGCTACCCGAGGTGAAGGCG 730
184 LeuMetPheArgIleAlaMetArgIleLeuLeuGlyCysGluProGlyPr 200
731 CTCATGTTCCGATCGCATGCGCATCTGCTGCGGCTCGAGCGCGGTCC 780
200 ALAGLysIleGlyLysAspGluGlnGlnLeuValGluAlaPheGluGln 217
781 AGCGGCGCGCGGAGAGAGACGACGACACTCTGTGAGGCTTTCCAGAGCA 830
217 eThrArgAsnLeuPheSerLeuProIleAspValProPheSerGlyLeu 233
831 TGACCGCAATCTCTTCTCTCCATTTAGCAGTTCAGGCGCTG 880
234 TyrArg..... 235

881 TACCGGGTAAGGGCGTTTCCGAGTCGAGTAGGGGCAACCGAAGCTCGG 930
236Gly.ValLysAlaArg 240
931 GCATCCGCTACCGCCACGCTCTCTCCGCGCTCAAGGGGTGAAAGCGCGG 980
241 AsnLeuIleHisAlaArgIleGluGlnAsnIleArgAlaLysIleArg 257
|||||
981 AACCTTATACAGCGCGCGCATCGAGAGAACATTCGCGCAAGATCCCGCG 1030
257 GluGlnIleThrGluProAspGlyGlyCysLysAspAlaLeuGlnLeu 274
|||||
1031 GCTTAGGCTACAGAGCCGATGGGGTTTGCAGAGGACGCGCTGCAGCTCC 1080
274 euIleGlnHisSerTrpGluArgGlyLysArgLeuAspMetGln..... 288
1081 TGATTGAGACTCGTGGGAGAGGGGAGAGAGCTGATATGACAGGTGAGA 1130
288 288
1131 AGCAATTTCAAAGGTGCCAAGGGCGGGGAGTGCCTGTGACTTTCAGA 1180
288 288
1181 CACACTTTCGGGGTCTCCAAAGCCCTGTCAAGGCCCAAGCTACTTCCA 1230
288 288
1231 GTGGGGGGGATGTAGTCTAGCTTTCAACTGTGGGTGCTGACCC 1280
288 288
1281 CTTCAAGGACCAACACCTTTGAGAGGGTGCCTAAGACATCTGC 1330
288 288
1331 AATCGCATTTTACATCAAGAAACATAACATAGCAAAATTACCGTTAT 1380
288 288
1381 GAAGTAGCAACAAGATAATTTATCGTTGGGGTCCACCACAACAGAG 1430
288 288
1431 AACCGATTAAAGGTGGCATTTGCTAGAGACTGTGGAAGGGGTGCG 1480
288 288
1481 TGAGCAATGGGGAAGATCCCAAGTTCAAAAGGCAAGGCTCATTCACAA 1530
288 288
1531 GGTAAAGGGAAGAGAGGATTAAAGGAGTTTGGCTTTTGTGTGG 1580
288 288
1581 TCTTTACTTTTATGACAAAACGATTTTACCTTGAAGCTTCCGCG 1630
288 288
1631 CAATATTTCAGGTGAGTCTTTGTAACAGTGTATAAATGACACTCAGA 1680
289AlaLeuLysGlnSerSerThr 295
1681 TCTGTATAAAGCTTCGTTTATCTTAACTTAAACATCTGTCACAA 1730
296 GluLeuLeuPheGlyLysGluThrThrAlaSerAlaAlaThrSer 312
|||||
1731 GAGCTCTCTTTGGTGTATGAACACTACAGCCAGTGTGCGAGCTCACT 1780
312 uIleThrTyrLeuGlyLeuTyrProHisValLeuGlnLysValArgGlu 329
|||||
1781 GATCACTTACCTAGACTCTACCCACATGTCTCCAGAAAGTTGAGAG 1830

329 IuIleLysSer..... 332
1831 AGATTAAAGACCAAGTAGATGATTCTAGAGTTCCCATTTGCTTAGGA 1880
332 332
1881 CATTCCTCTATTAAACACACACACACACACCCCACTGTATATAATTGCTC 1930
332 332
1931 GATACACCACTACTATAGACATGAAGATCTGAGAGTACGTGGAGCTG 1980
332 332
1981 GGGGAGAGACTCCACTCGTAATTTAAAGGACAGTTGTTGACTGG 2030
332 332
2031 GCTCTCTTTGGGAGAAATTGACCCCTCTCTCTCTCTCTCTCTCT 2080
332 332
2081 CCTCTCTCTCTCCACACACACACATCAACACCTTTATAGACAAAG 2130
332 332
2131 TTTCTCTTCCCTGACCAAGACATGAATATGATTAGACCAATAGC 2180
332 332
2181 TGATCAGGTGCGCAGTGTGTGAGGCTCAGGGTATGACCTTATATA 2230
332 332
2231 CCGTATAGCAACATGTCTGATTAATGGTTTAGCTGAGGAAGTGG 2280
332 332
2281 AAAGGAAGCCATCAGGCATCAGCTCTTCCCTTATCTCCATC 2330
332 332
2331 CAGAGCCTTCAGTTTATGTAACAGTGAAGTCTCGGGGCTGACTTT 2380
333Lys. 333
2381 TTTTGGAGTCCCGAGGATCCATCACTCTTTTATCTGTTCCATA 2430
334 .GlyLeuLeuCysLysSerAsnGlnAspAsnLysLeuAspMetGluThrL 350
|||||
2431 GGCTTACTTTGCAAGAGCATCAAGACACACAGTTAGACATGGAACCT 2480
350 euGlnGlnLeuLysTyrIleGlyCysValIleLysGluThrLeuArgLeu 366
|||||
2481 TGGCACAGCTTAATAATACACTGGGTGTCTCATTAAGGACACCTGGATG 2530
367 AsnProProValProGlyGlyPheArgValAlaLeuLysThrPheGlu 382
|||||
2531 AATCTCTCGGTTCCAGAGGTTTGGGGTGTCTCTGAAGACTTTTGAGCT 2580
382 382
2581 GAATGTGAGTGACCTCGTGTCCCAACCCCAAGCCTGTGCAAGTCA 2630
382 382
2631 CTCTCTATGCTGTGACATCAGCTGCCAGAGAGTGGCTCACTGCC 2680
382 382
2681 TTGACAGTGTCTCGCTCTATGTAAGTACGGGAACCAATTTGCTCTCTCT 2730


```

||||| ||||| ::::::::::::::::::::
588 GCTGCAAAAAGACTCTGC.....GTGCTGGTTATCCAGAAA 625
181 allysArgLeuMetPheArg11LeuMetArg11LeuLeuGlyCysGlu 197
626 TGAAGAACTCATGTTTGGATTACCTATGAGAAATCTGCTGTTGTTGAA 675
198 ProGlyProAlaGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 214
676 CCAGAGCAAAATA.....AAGAGCGAGAGAGAACTGGTGGAAAGCTTT 719
214 eGluGluMetThrArgAsnLeuPheSerLeuProIleAspValProPheS 231
720 TGAGGAATATGATCAAAAACCTGTTCTCTCCCAATGACAGCTTCCTTCA 769
231 eArg1LeuThrArgGlyValLysAlaArgAsnLeuIleHisAlaArg1Le 247
770 GTGGCTGTACAGGGGTTTGAGGCGACCAATTTCACTCACTCCAAATTT 819
248 GluGluAsnIleArgAlaLysIleArgArgLeuGlnAlaThrGluProAs 264
820 GAGGAAACATCATCAGAGAAATTCAGAT...GACGACAATGAAAAACGA 866
264 pGlyGlyCysLysAspAlaLeuGlnLeuIleGluHisSerTrpGluA 281
867 ACAGAAATACAAAGAGCCCTTCAGCTGTGATGAGAAACGAGAAAGAA 916
281 rGlyGlyArgLeuAspMetGlnAlaLeuLysGlnSerSerThrGluLeu 297
917 GTGACGAACCTTTAGTTTGCAGCGATGAAAGAAAGCAGCTACAGAGCTT 966
298 LeupheGlyGlyHisGlyThrThrAlaSerAlaAlaThrSerLeuIleTh 314
967 CTATTGTGAGGTCAATGAAACACGCGCCAGCATGCAACCTCACTGTCTAT 1016
314 rTyrtLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLe 331
1017 GTTCTGGGTGGAACACAGAGTGTGTCAGAGAGTCCAGAGAGAGGTTTC 1066
331 ySerLeuGlyLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 347
1067 AGGAGAAAGTTGAAATGGGCAATGTATCACTGGAAGAGGCTTGATGATG 1116
348 GluThrLeuGluGlnLeuLysTyrtIleGlyCysValIleLysGluThrLe 364
1117 GAGCGTGTGGACACAGCTGATGATGATGATGATGATGATGATGATGATG 1166
364 uArgLeuAsnProproValProGlyGlyPheArgValAlaLeuLysThrP 381
1167 TGAATTCAAACCTCTCTGTTCCCGGAGGATTCAGACTCCGACACAAACCT 1216
381 heGluLeuAsnGlyTyrtGlnIleProLysGlyTrpAsnValIleTyrtSer 397
1217 TGAATTCAAATGATGATGATGATGATGATGATGATGATGATGATGATG 1266
398 IleCysAspThrHisAspValAlaAspIlePheThrAsnLysGluGluPh 414
1267 ATCTGTGACACGACGATGTGCGACGCTTTTCCAAACAAAGAGAGAGTT 1316
414 eAsnProAspArgPheIleValProHisProGluLysAspAlaSerArgPheS 431
1317 CCAGCGGAGAGATTCATGAGCAAGAGTGTGAGAGCGGCTCCAGATTTA 1366
431 ePheIlePheProheGlyGlyGlyLeuArgSerCysValGlyLysGluPhe 447
1367 ACTACATCCCTTCGAGAGAGATCCAGAGATGTGTGTGGCAAAAGATTTC 1416
448 AlaLysIleLeuLeuLysIlePheThrValGluLeuAlaArgHisCysAs 464
1417 GCCAAAGTGTACTCAAGATCTTTTAACTTGAAGTTAACCCAGCATTTGCA 1466
464 pTrpGlnLeuLeuAsnGlyProProThrMetLysThrSerProThrValT 481
::: ||| ::::::::::::::::::::

```

```

1467 TTGATTCTCTCAAGAGCAACCCCGACATGAAACAGGCCGACTATT 1516
481 yProValAspAsnLeuProAlaArgPheThrTyrtPhe 493
1517 ACCCACTGGACATCTCCCTACCAATTCATGATGAT 1554
seq_name: /SIDB1/gcdata/geneseq/geneseqn/NA1998.DAT:AAV12203
seq_documentation_block:
ID AAV12203 standard; cDNA; 1850 BP.
XX
AC AAV12203;
XX
DT 22-JUN-1998 (first entry)
XX
DE zebrafish retinoid metabolizing protein zp450RAI cDNA.
XX
KW Retinoid metabolizing protein; P450RAI; retinoid oxidase;
KW retinoic acid; zebrafish; inhibitor; antisense; cancer;
KW actinic keratosis; oral leukoplakia; head tumour; neck tumour;
KW non-small cell lung carcinoma; basal cell carcinoma;
KW acute promyelocytic leukaemia; skin cancer; acne; psoriasis;
KW ichthyosis; therapy; diagnosis; screening; ss.
XX
OS Danio rerio.
XX
FH Key Location/Qualifiers
FT CDS 88..1566
FT /tag= a
XX
PN MO9749815-A1.
XX
PD 31-DEC-1997.
XX
PE 23-JUN-1997; 97WO-CA00440.
XX
PR 01-OCT-1996; 96US-0724466.
PR 21-JUN-1996; 96US-0667546.
XX
PA (TOOH ) UNIV QUEENS KINGSTON.
XX
PI Beckett BR, Jones G, Petkovich PM, White JA;
XX
DR MPI: 1998-077178/07.
XX
DR P-PSDB; AAM12203.
XX
PT Retinoid metabolizing protein - useful to develop products to treat,
PT e.g. cancer, actinic keratosis, oral leukoplakia, acne, psoriasis or
PT ichthyosis
XX
PS Claim 15; Page 52-54; 110pp; English.
XX
CC This cDNA clone codes for a novel zebrafish retinoid metabolizing
CC protein (see AAM44159) designated zp450RAI. The encoded protein is a
CC retinoid oxidase that has the ability to hydroxylate retinoic acid
CC at the 4 position of the beta-ionone ring and is inducible in
CC epithelial cells exposed to retinoic acid. To isolate the clone,
CC transcripts present in fin tissue regenerating in the presence or
CC absence of retinoic acid were compared using a differential display
CC PCR technique, and an isolated clone (see AAV12213) was used to screen
CC a D. rerio 6-18 hr embryo cDNA library. Human (see AAV12204) and
CC mouse (see AAV12205) P450RAI clones have also been identified. The
CC isolated clones can be used in the recombinant production of
CC P450RAI proteins. Antisense nucleic acids can be used in a claimed
CC method for inhibiting retinoic acid hydroxylation for the treatment
CC of cancer, actinic keratosis, oral leukoplakia, secondary tumours of
CC the head and/or neck, non-small cell lung carcinomas, basal cell
CC carcinomas, acute promyelocytic leukaemia, skin cancer, and
CC premalignancy associated with actinic keratosis, acne, psoriasis
CC and/or ichthyosis. Promoter sequences (see AAV12206-08) are also
CC claimed.
XX
SQ Sequence 1850 BP; 519 A; 391 C; 454 G; 486 T; 0 other:

```

alignment_scores:
 Quality: 1734.00 Length: 496
 Ratio: 4.070 Gaps: 4
 Percent Similarity: 85.887 Percent Identity: 67.742

alignment_block:
 US-09-668-482-32 x AAV12203 ..

Align seg 1/1 to: AAV12203 from: 1 to: 1850

```

1 MetGlyLeuProAlaLeuAlaSerAlaLeuCysThrPheValLeu 17
  ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
88 ATGGGGCTGTACACCTTATGGTCACTTCTCTCCACCATGCTGCTACC 137
17 OleuLeuPheLeuAlaLeuLysLeuThrPaspLeuTyrCysVal 34
  ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
138 CGTTTACTCTTTCGCCGCCGGTGAAGTTGGAGATGTTAATGATCC 187
34 erSerArgPaspArgSerCysAlaLeuProLeuProGlyThrMetGly 50
  ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
188 GACGAGTGATCCGAAGTCAAGAACTCCCTACCCGCAAGTACATGGGC 237
51 PheProPhePheGlyGlyThrLeuGlnMetValLeuGlnArgArgLysPh 67
  :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
238 TTGCGCTTCATTGGAGAAACGCTCCAGCTGATCCACAGAGAAAGATT 287
67 eLeuGlnMetLysArgArgLysTyrGlyPheIleTyrLysThrHisLeu 84
  ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
288 TCTGGCGATTGAAGCGCAAAATACGGGTGCATCTACAAAGACGACCTCT 337
84 heGlyArgProThrValArgValMetGlyAlaAspAsnValArgArgLys 100
  ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
338 TCGGGAACCCGACTGTCAGGGATGAGAGCTGATATGTGAGGACAGATT 387
101 LeuLeuGlyGlnHisArgLeuValSerValHisThrProAlaSerValArg 117
  ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
388 CTGCTGGGCGACACAAAGCTGTCTGTCTGTCTGACGACATCAGTAG 437
117 gThrIleLeuGlnLysAlaGlyCysLeuSerAsnLeuHisAspSerHisSL 134
  ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
438 AACCATCTGGGCTGTGACACCTCTCCATGTTCATGAGATTCAACACA 487
134 ysgLnaArgLysLysValIleMetGlnAlaPheSerArgGluAlaLeuGln 150
  ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
488 AAAACAAGAAAAAGCCATTATGAGGGCTTCTCGAGATGCTCTGGAG 537
151 CysThrValLeuValIleAlaGlnGluValSerSerCysLeuGlnIleTyr 167
  ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
538 CACTACATTCCTCCGTGATCCAGCAGAGGTGAAGCCCATATACAGAAATG 587
167 pLeu.....SerCysGlyGluArgGlyLeuLeuValTyrProGly 181
  ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
588 GCGSCAAAAAGACTCTGCTG.....GTCTGTGTTTATCCAGAA 625
181 allysArgLeuMetPheArgIleAlaMetArgIleLeuLeuGlyCysGlu 197
  :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
626 TGAAGAAATCATGTTTCGATACCTATGAGAATCTGCTGTTTGAA 675
198 ProGlyProAlaGlyGlyGluAspGlnGlnIleValGluAlaPh 214
  ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
676 CCAAGACCAATA.....AAGACGAGCAGCAAGAACTGGTGGAGCTTT 719
214 eglGluMetThrArgAsnLeuPheSerLeuProIleAspValProPhe 231
  ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
720 TGAAGAAATGATCAAAACTGTCTCTCTTCCCATGACATGACCTTCTTCA 769
231 erGlyLeuTyrArgGlyValLysAlaArgAsnLeuIleHisAlaArgIle 247
  ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
770 GTGCTGTGTACAGGGGTTTGAGGGCAGCAGCAATTCATTCACCCAAATTT 819
248 gLIGluAsnIleArgAlaLysIleArgArgLeuGlnAlaThrGluProAs 264

```

```

820 GAGAAAAACATCGAGAAAGAAATTCAGAT...GACGACATGAAAAACA 866
264 pGlyGlyCysLysAspAlaLeuGlnLeuLeuIleGlnHisSerThrPglua 281
  :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
867 ACAGAAATACAAAGACGCCCTTCAGCTGTGATCGAAGACAGAGAA 916
281 rGlyGlyLysArgMetGlnAlaLeuLysGlnSerSerThrGluLeu 297
  ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
917 GTGACGAACTTTAGTTTCAGGCGATGAAGAAACACCTACAGAGCTT 966
298 LeuPheGlyGlyHisGlnThrThrAlaSerAlaAlaThrSerLeuLeu 314
  ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
967 CATATTGGAGGTGATCAAAACCAACCCGACACTGCAACTCATTGTCTAT 1016
314 rTyrLeuGlyLeuTyrProHisValLeuGlnLysValArgGluIle 331
  :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1017 GTTCTGGGCTGAAACACAGAAAGTGTGCAAGAGTCAGAGAGAGTTTC 1066
331 ySerLysGlyLeuLeuCysLysSerAsnGlnAspAsnLysLeuSpmet 347
  ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1067 AGGAGAGGTTGAAATGGGATGATATACCTGAAAGGGCTTGTGATAG 1116
348 GluThrLeuGlnGlnLeuLysTyrIleGlyCysValIleLysGluThrLe 364
  ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1117 GAGCTGTTGAGACACTGAAAGTACACTGATGTGTATTAAGACACTCT 1166
364 uArgLeuAsnProProValProGlyGlyPheArgValAlaLeuLysThrP 381
  ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1167 TAGAATCAACCTCCTGTTCCGGAGAGATTGAGATCGCATCAAAACCT 1216
381 heGluLeuAsnGlyTyrGlnIleProLysGlyThrPaspValIleTyrSer 397
  ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1217 TTGAATTGATGCTTACCAAAATCTCTAAAGATGGAAGCTCTTTAAAGC 1266
398 IleCysAspThrHisAspValAlaAspIlePheThrAsnLysGluLysPh 414
  ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1267 ATCTGTGACAGCAGCAGATGTGGCCGACGCTTCCAAACAAAGACAGATT 1316
414 eAsnProAspArgPheIleValProHisProGluAspAlaSerArgPhe 431
  ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1317 CCAGCCGAGAGATTTCATGAGCAAAAGCTGTGGAGACGCGGTCCAGTTTA 1366
431 erPheIleProPheGlyGlyLeuArgSerCysValGlyLysGluPhe 447
  :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1367 ACTACATCCCTTCGGAGAGAGATCCAGAGATGTGTGTGGCAAGAGTTTC 1416
448 AlaLysIleLeuLeuLysIlePheThrValGluLeuAlaArgHisCysAs 464
  ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1417 GCCAAAGTGTACTCAAGATCTTTTACTTACTGATTAAAGCAGCATTCGAA 1466
464 pTyrGlnLeuLeuAsnGlyProProThrMetLysThrSerProThrValT 481
  ||| ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1467 TTGGATTCTCTCAAAAGCAGACCCCGACAAATGAAAAACGGCCGACTATT 1516
481 yProValAspAsnLeuProAlaArgPheThrTyrPhe 493
  ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1517 ACCCAGTGGACAATCTCCCTTACCAAAATTCACCTAGTTAT 1554

```

seq_name: /SIDSI/gcgdata/geneseq/geneseqn/NA1998.DAT.AAV12214

seq_documentation_block:
 ID AAV12214 standard; DNA; 2677 BP.

XX AAV12214;
 XX
 XX 22-JUN-1998 (first entry)
 XX
 XX Human retinoid metabolizing protein hp450RA1 genomic DNA.
 DE Retinoid metabolizing protein; P450RA1; retinoid oxidase;
 KW Retinoic acid; human; inhibitor; antisense; cancer;
 KW actinic keratosis; oral leukoplakia; head tumour; neck tumour;
 KW

KM non-small cell lung carcinoma; basal cell carcinoma;
 KW acute promyelocytic leukaemia; skin cancer; acne; psoriasis;
 KM ichthyosis; therapy; diagnosis; screening; cytochrome P450; ds.
 XX
 OS Homo sapiens.

XX WO9749815-A1.

XX 31-DEC-1997.

XX 23-JUN-1997; 97MO-CA00440.

XX 01-OCT-1996; 96US-0724466.

XX 21-JUN-1996; 96US-0667546.

XX (TOOH) UNIV QUEENS KINGSTON.

XX Beckett BR, Jones G, Petkovich PM, White JA;
 WPI; 1998-077178/07.

XX Retinoid metabolising protein - useful to develop products to treat,
 PT e.g. cancer, actinic keratosis, oral leukoplakia, acne, psoriasis or
 PT ichthyosis

PS Disclosure; Page 68; 110pp; English.

XX This sequence comprises a genomic clone of a novel human retinoid
 CC metabolising protein designated hp450RAI. Another genomic clone
 CC (see AAV12215) is also provided, as well as a cDNA clone (see AAV12204)
 CC that includes a coding sequence for the full-length hp450RAI
 CC protein (see AAV44160). The encoded protein is a retinoid oxidase
 CC that has the ability to hydroxylate retinoic acid at the 4 position
 CC of the beta-ionone ring and is inducible in epithelial cells
 CC exposed to retinoic acid. The hp450RAI gene has been localised to
 CC 10q23-24. Zebrafish, human and mouse P450RAI cDNA clones (see
 CC AAV12203-05) are claimed. They can be used in the recombinant
 CC production of P450RAI. Antisense nucleic acids can be used in a
 CC claimed method for inhibiting retinoic acid hydroxylation for the
 CC treatment of cancer, actinic keratosis, oral leukoplakia, secondary
 CC tumours of the head and/or neck, non-small cell lung carcinomas,
 CC basal cell carcinomas, acute promyelocytic leukaemia, skin cancer,
 CC and premalignancy associated with actinic keratosis, acne,
 CC psoriasis and/or ichthyosis. Promoter sequences (see AAV12206-08)
 CC for the P450RAI genes are also claimed.

XX Sequence 2677 BP; 489 A; 775 C; 868 G; 511 T; 34 other;

alignment_scores:

Quality: 1251.00 Length: 669
 Ratio: 4.075 Gaps: 7
 Percent Similarity: 45.889 Percent Identity: 43.797

alignment_block:

US-09-668-482-32 x AAV12214 ..

Align seg 1/1 to: AAV12214 from: 1 to: 2677

1 MetGlyLeuProAlaLeuAlaSerAlaLeuCysThrPheValIleuPr 17
 676 ATGGGGCTCCCGCGCTGCTGGCCAGTGGCTCTGCACCTTGCTGCTGCC 725
 17 oleuLeuLeuPheLeuAlaAlaLeuLysLeuTrpAspLeuTYrCysValS 34
 726 GCTGCTGCTCTCTCTGCTGCATCAAGCTCTGGGACCTGCTACTCGTGA 775
 34 eTserTrgAspArgSerCysAlaLeuProLeuProGlyThrMetGly 50
 776 GCGGCGGACGCGCAGTGTGCCCTCCCAATGCCCCCGGAGCTATGGAS 825
 51 PheProPhePheGlyGlyLthrLeuGlnMetValLeu.Gln..... 63

826 TTCCCTCTTTGGGAAACSTTGCAGATGNTACTNCAAGTAAGGAGAGG 875
 63 63
 876 TGGGGCGGAGACAGGCTGTCCCGGAGCCGGCGGCTGTGGCTTCT 925
 63 63
 926 GCTGAAGTCGGGTAGCGCCCCCGGAGCATGCTATTGGCGCTAGAG 975
 63 63
 976 CAGGCTGGCGGAGCGCGGCTCCCGMKYMSCATGACGSCRMWK 1025
 64 ArgArgLysPheLeuGlnIleLysArg 72
 1026 TMMCTCCGCTTTCCTCCAMAGCGARSARNKCYKMRGATGAAGCGC 1075
 73 ArgLysTYrGlyPheIleTYrLysThrHisLeuPheGlyArgProThrVa 89
 1076 AGGAATACGGCTTTCATCTCAAGACGCACTGTGGGGGCCACCGT 1125
 89 IArgValMetGlyAlaAspAsnValArgArgIleLeuLeuGlyLHisA 106
 1126 ACGGTGATGGCGCGGACAAATGTGGCGCATCTGTGAGACACC 1175
 106 rGlyValIserValHisTrpProAlaSerValArgThrIleLeuGlyAla 122
 1176 GCGGTGTGTGCTGCTCACTGGCAGCGTGGTGGCAGCATCTGGGATCT 1225
 123 GlyLysLeuSerAsnLeuHisAspSerSerHisLysGlnArgLysLys... 138
 1226 GCGTCTCTTCACTCACTGACAGCATCTCTCCACAGAGCGCAAGAA 1275
 138 138
 1276 GGGGCGAGAGCGAGCGCTGACAGGAGGAGGCCCATTTATGAGCG 1325
 138 138
 1326 GAATTCGGCTGATGATGCTAGCGCGGCTAGCAGCTTGAGTGGGCT 1375
 138 138
 1376 AGGACCTTGCCAGCTCCAGGTTAGCTTCCACAGCTCGGAGAGTGCCAT 1425
 138 138
 1426 GTGTCTGGCAGACTGGGGGTGTCTGGAAGGGGACGGGTAGACGAG 1475
 139 ValIle 140
 1476 GGGCGATGAGGCTTTTAACTGTGCTCCCTCGGAGACTCAGGTAT 1525
 141 MetGlnAlaPheSerArgGluAlaLeuGlnCysTYrValLeuValIleAl 157
 1526 ATGGGGCTTCAAGCCGAGGACGACCTGGAATGTACGCGCGGTATCAC 1575
 157 aGluGluValSerSerCysLeuGluGlnTrpLeuSerCysGlyLysArg 174
 1576 CGAGCAAGTGGCAGACCTGAGCAAGTGTGCTGAGCTGCGCGCAGGCG 1625
 174 LysLeuLeuValTYrProGluValLysArgLeuMetPheArgIleAlaMet 190
 1626 GCTCTCTGTCTTACCCAGAGTGAAGCGCTCATGTTCGAAATGGCATG 1675
 191 ArgIleLeuLeuGlyCysGluProGlyProAlaGlyGlyLysAspGly 207
 1676 CGCATCTTACTGGGCTGCGAACCACCACTGGCGGGCGACGGGACTCGGA 1725
 207 uGlnGlnLeuValGluAlaPheGluGlnMetThrArgAsnLeuPheSerL 224
 1726 GCAGCAGCTTGAGAGGCTTGAAGAAATGACCCGCAATCTTCTTCGCG 1775

224 euProIIleaspyAlPropheserGlyLeuTyrArg..... 235
|||||
1776 TCCCATCGACGTGCCCTTCACGGGGCTGTACCGGGTAGGGCGGCAAC 1825
235 235
1826 GGGGTGCGGACTAGGGGGCGGGACCTGGGGCTGTGCTCACGCCCGCGGG 1875
236Gly.ValIysAlaArgAsnLeuIleHisAlaArgIle 247
1876 CTCCTGCGCTGAGGGGCATGAGGGCGGCACTTCCTCAGCGCGCATC 1925
248 GluGlnAsnIleArgAlaIleArgIleArgIleGlnAlaThrGluProAs 264
|||||
1926 GAGCAGAACATTCGGCGCCAGATCTGCGGGCTGCGGGCATCCGAGCGGG 1975
264 pGlyGlyIysIysAspAlaLeuGlnLeuIleGluHisSerTrpGluA 281
|||||
1976 CCAGGGCTGCAAAAGACGGCTGCACGTGTGATCGAGCAGCTGCGGAGA 2025
281 rGlyGluArgLeuAspMetGln..... 288
2026 GGGGAGACGGCTGACATGACGTGAGTACGACCTTCAGACCGACGCACT 2075
288 288
2076 GCGGAGTTTGGTCCCTGGCTTTCACAGGCGGTGTCTCGGGGGCCCCAA 2125
288 288
2126 AGCGCGCGCTGGGGCCAGCTTCTGAGAGTGGCGCGCGCTCAGACTA 2175
288 288
2176 CAGTATGGAATCCCGAAGAAGGTGAGACACCGGTCAGAGAGCTGC 2225
288 288
2226 GGAAGGGGCTCGGGMGAACCTGGAGCATCCCTAGCCTTAMCAGGTT 2275
288 288
2276 TCAGAAGGAAGTGGAAATTGCAAAAATTGTAATAAGAACCTTGGCAT 2325
288 288
2326 TTTAATAAACTAGACTTTAACTCAGAGATTCCGGTAGGCGGGCTCG 2375
288 288
2376 TACTCGCTTACTGCTCAGCTGACTAAGGAGGACGTTGATTTGTTTA 2425
288 288
2426 AAGATATTGTTTCTTCTGACTTCTGTACGAAAAACATTTAGCCCTCTA 2475
288 288
2476 GTCTTCCCTCCAGAACTCTCAGTTCGATCTGAGTAATCTCTGTCAAA 2525
289Ala..Leu 290
2526 CCGCAGGCACTTGTGAGATGTGGGTCTCACTCTATCTTACGACCTA 2575
291 LysGlnSerSerThrGluLeuLeuPheGlyIysGlnThrThrAla 307
|||||
2576 AAGCAATCTTCAACCGAAGCTCCCTTTGGAGACACGAAACCGAGCGCAG 2625
307 rAlaAlaThrSerLeuIleThrTyrLeuGlyLeuTyrProHisValLeu 324
|||||
2626 TGCAGCCACATCTCTGTGATCACTTACCTGGGGCTCTACCCACATGTTCTCC 2675

324 ln 324
||
2676 AG 2677
seq_name: /Sids1/gcgdata/geneseq/geneseqn/NA2000.DAT: AAC39029
seq_documentation_block:
ID AAC39029 standard; DNA, 1685 BP.
XX AAC39029;
XX 17-Oct-2000 (first entry)
XX
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 23101.
DE
XX Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
XX Arabidopsis thaliana.
XX
XX EPI033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 990S-0121825.
XX 05-MAR-1999; 990S-0123180.
XX 09-MAR-1999; 990S-0123548.
XX 23-MAR-1999; 990S-0125788.
XX 25-MAR-1999; 990S-0126264.
XX 29-MAR-1999; 990S-0126785.
XX 01-APR-1999; 990S-0127462.
XX 06-APR-1999; 990S-0128234.
XX 08-APR-1999; 990S-0128714.
XX 16-APR-1999; 990S-0129845.
XX 19-APR-1999; 990S-0130077.
XX 21-APR-1999; 990S-0130049.
XX 23-APR-1999; 990S-0130510.
XX 28-APR-1999; 990S-0130891.
XX 30-APR-1999; 990S-0131449.
XX 30-APR-1999; 990S-0132048.
XX 04-MAY-1999; 990S-0132407.
XX 05-MAY-1999; 990S-0132484.
XX 06-MAY-1999; 990S-0132485.
XX 06-MAY-1999; 990S-0132486.
XX 07-MAY-1999; 990S-0132487.
XX 11-MAY-1999; 990S-0132863.
XX 14-MAY-1999; 990S-0134218.
XX 14-MAY-1999; 990S-0134219.
XX 14-MAY-1999; 990S-0134221.
XX 18-MAY-1999; 990S-0134370.
XX 18-MAY-1999; 990S-0134768.
XX 19-MAY-1999; 990S-0134941.
XX 20-MAY-1999; 990S-0135124.
XX 21-MAY-1999; 990S-0135353.
XX 24-MAY-1999; 990S-0135629.
XX 25-MAY-1999; 990S-0136021.
XX 27-MAY-1999; 990S-0136392.
XX 28-MAY-1999; 990S-0136782.
XX 01-JUN-1999; 990S-0137222.
XX 03-JUN-1999; 990S-0137528.
XX 04-JUN-1999; 990S-0137502.
XX 07-JUN-1999; 990S-0137724.
XX 08-JUN-1999; 990S-0138094.
XX 10-JUN-1999; 990S-0138540.
XX 10-JUN-1999; 990S-0138847.
XX 14-JUN-1999; 990S-0139119.
XX 16-JUN-1999; 990S-0139452.
XX 16-JUN-1999; 990S-0139453.
XX 17-JUN-1999; 990S-0139492.

PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 09-JUL-1999; 99US-0142803.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 10-AUG-1999; 99US-0147935.
PR 11-AUG-1999; 99US-0148171.
PR 12-AUG-1999; 99US-0148319.
PR 13-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.

PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 13-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

alignment_scores: Quality: 474.00 Length: 491
 Ratio: 1.612 Gaps: 12
Percent Similarity: 59.878 Percent Identity: 27.291

alignment_block:
US-09-668-482-32 x AAC39029 ..

Align seg 1/1 to: AAC39029 from: 1 to: 1685

15 ValueProLeuLeuLeuPheLeuAlaLeuLeuYsLeuTriPheLeuTy 31

PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 18-JUN-1999; 99US-0139817.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140635.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143547.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.

PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148655.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.

```

PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

alignment_scores: quality: 455.50 length: 486
 ratio: 1.627 gaps: 13
Percent Similarity: 57.613 Percent Identity: 28.189

alignment_block:
US-09-668-482-32 x AAC48714 ..

Align seg 1/1 to: AAC48714 from: 1 to: 1398

```

1 MetGlyLeuProAlaLeuAlaSerAlaLeuGlyThrPheValLeuPr 17
2 ATGGGCTCTCTCTTGATCATCGGTCT...CTGTGT..... 54
17 OleuLeuLeuPheLeuAlaLeuLysLeuTrpAspLeuTyrCysVal 34
22 ATGGGCTCTCTCTTGATCATCGGTCT...CTGTGT..... 54
34 eSerAlaGAspArgSerCysAlaLeuProLeuProProGlyThMetGly 50
73 ..AATCGATGCGGATATACCAAGAAATGATCTCTCCGGAACCATGGGC 120
51 PheProPhePheGlyGluThrLeuGlnMetValLeuGlnArgArgLysP 67
121 TGGGCATCTTTGGCGAAACACCGAGATTCTCAAAACAGGCCCAACTT 170
67 eLeuGlnMetLysArgArgLysTyrGlyPheLeuTyrLysThrHisLeu 84
171 CATGAGAAACCAAGACTCCGATACGGAGAGTTCTTCAAAATCTCATCTTC 220
84 heGlyArgProThrValArgValMetGlyAlaAspAsnValArgArgL 100
221 TAGGTGTCCAGAGTTATCTCAATGGAGTCAGAAAGTAACAGATACATT 270
101 LeuLeuGlyGluHisArgLeuValSerValHisTrpProAlaSerValAr 117
271 TTAAAGAACCAATCAAAAGTTTGGTCCCGGTACACCAATGATGACT 320
117 gTrnIleLeuGlyAlaGlyLeuSerAsnLeuHisAspSerSerHisL 134
321 TGATATACCTTGGGACTTGTAACATGGCTGCGGTTCAACGGTTCCAGCCAC 370

```

```

134 ySGLnArgLysLysValIleMetGlnAlaPheSerArgLysValLeuGln 150
371 GGCTTATGAGAGCTCGCTTGTCTCTCATAGCTGACCATGATGAGAA 420
151 CysTyrValLeuValIleAlaGluGlu...ValSerSerCysLeuGlu 166
421 GATCATATCTTGCCTAAAGTTGATCATCTTCATGACAGAGCTATCTTGATCA 470
166 nTrpLeuSerCysGlyGluArgGlyLeuLeu...ValTyrProGluValL 182
471 GTGG.....AATGACCTTGAGTTATGATATCCAAAGATTAAGACCA 511
182 ySArgLeuMetPheArgLysIleAlaMetArgLysLeuGlnLysGluPro 198
512 AACATATGTCATTTTATATCTTCACTGACACAAATGCTGGAGACTTAAGA 561
199 GlyProAlaGlyGlyGlyLysAspGluGlnGlnLeuValGluAlaPheG 215
562 AAACCATTT.....GTGAGAAATTCAAAACTGCATTTCTT 596
215 uGlnMetThrArgAsnLeuPheSerLeuProIleAspValProPheSerG 232
597 CAAGCTTGTTGTGGAGCTTATCCGTTCCGATGATCTCCGGGCACAA 646
232 LLeuTyrArgGlyValLysAlaArgAsnLeuIleHisAlaArgLysL 248
647 ATTATCTTGCGGAAATCCAAAGCAAGAAAT..... 675
249 GluAsnIleArgAlaLysIleArgArgLeuGlnAlaThrGluProAsp 265
676 ..AACATTTGATAGGTGCTTAAGAGAGCTGATGCAAGACGTAGAGATTTC 722
265 yGly.....CysLysAspAlaLeuGlnLeuLeuIleGluHisSerTrpG 280
723 TGGAGAAACATTTACAGACATGTTAGTTACTGATGAAGAGAGAGTGA 772
280 LuArgGlyGluArgLeuAspMetGlnAlaLeuLysGlnSerSerThrGlu 296
773 ACCGATACCCGTTAACCGAT...GAAGAGATTAAGAGACCAAGTTGTGACG 819
297 LeuLeuPheGlyGlyHisGluThrThrAlaSerAlaAlaThrSerLeu 313
820 ATTGTGTATTCGGGTTACGAAACGTCTCTACGACCTCAATGATGCGCTCT 869
313 eThrTyrLeuGlnLeuTyrProHisValLeuGlnLysValArgLysL 330
870 TAACTACCTTATGATGACCCAAAGACTTTCAGAGACTGAGAGCTGAG. 918
330 LeuLysSerLysGlyLeuLeuCysLysSerAsnGlnAspAsnLysLeuAsp 346
919 .....CATTTGCAATTCAGGAAAGAAAGAAACAGAGACCAACCACTGCGT 963
347 MetGluThrLeuGlnGluGlnLeuLysTyrIleGlyCysValIleLysGlu 363
964 CTTCAGAGACGTGAAGTCATAGAAATGATCACTGACGCTGATTTATGAGAC 1013
363 rLeuAlaGlyLeuAsnProProValProGlyGlyPheArgValAlaLeuLys 380
1014 ATCAAGATTTGGCAACAGATCGTTATGGGTTCTTAAGGAAACTACTGCTG 1063
380 hrPheGluLeuAsnGlyTyrGlnIleProLysGlyTyrAsnValIleTyr 396
1064 ACTTGAGAAATCAACGTTATTTAATCCCAAAAGATGAGACAAATTAAGTA 1113
397 SerIleCysAspThrHisAspValAlaAspIlePhehrAsnLysGlu 413
1114 TACACGAGCGAAATTAATTAACGATGCAAAATCTTTATGAGAGCCCATTTGAT 1163
413 upHeAsnProAspArgPheIleValProHisProGluAspAlaSerArgP 430
1164 CTTTAATTCATGAGATGATGAAGAGAGCTTGAG.....TCACAA 1207
430 heSerPheIleProPheGlyGlyLeuArgSerCysValGlyLysGlu 446

```



```
732 GAGCGCTGACGCTGGTGTGATGAAAAGAGAGAGAGAGAGAGAG 781
265 YGLY.....CysLysAspAlaLeuClneule.....GluHis 278
782 ACCGGAGAGAAAGATATGCTTGGCGGCTTGGCGGAGGAG 831
278 eTTPGLuArgLysLysLeuAspMetGlnAlaLeuLysGlnSer 294
832 GATTTTCGATGAGAGATGTTGACTTC.....TTG 863
295 ThrGluLeuLeuPheGlyGlyHisGluThrThrAlaSerAlaAlaThrSe 311
864 GTGGCTTACCTTGTGCGCGTATGAAACACCTCCAGCATGACTCT 913
311 rLeuLethrThrLysGlyLeuThrProHisValLeuGlnLysValArg 328
914 CGCCGTCAAATTTCTACCCGAGACTCTTACCTCTGCTGCTCAACG 963
328 LuGluLysSerLysLysLeuLysSerAsnGlnAspAsnLys 344
964 AAGAG.....CATGAAAGATTAGGCAATGAGAGATGATCGTACT 1007
345 LeuAspMetGluThrLeuGluGlnLeuLysThrLysGlyCysValLys 361
1008 CTGTGATGAGATGATTACAGTCATGCTTACACCAATGTTGTTAA 1057
361 sGluThrLeuArgLeuAsnProProValProGlyGlyPheArgValAla 378
1058 TGAGACGCTAGAGAGGCTTACATCATCGGGGCTTTTCAGACTGCA 1107
378 euLysThrPheGluLeuAsnGlyTyrglnLysProLysGlyTrpAsnVal 394
1108 TGACGAGATGTGAGATCAAAAGTTATAAATTCAAAAGGTTGAAAGTA 1157
395 lLeyrSerIleCysAspThrHisAspValAlaAspIlePheThrAsnLys 411
1158 TTCATCATGCTTACAGCGCTTCAATTAGACCCAAACCACTTCAAGATGC 1207
411 sGluGluPheAsnProAspArgPhe.....IleValProHisP 424
1208 TCGCAGCTTTCACACCTTGGAGATGGCAGACGACTCGTAAAGCAGAGCC 1257
424 roGluAspAlaSerArgPheSerPheIleProPheGlyGlyLeuArg 440
1258 CTTCCTAATGTG.....TTCACACCGTTTGGTGAGGCGCAAGG 1295
441 SerCysValGlyLysGluPheAlaLysIleLeuLysIlePheThrVal 457
1296 CTATGTCGCCGTTACGAGCTGGCTAGGGTTGCACTCTCTGTTTCCTTCA 1345
457 lGluLeuAlaArgHisCysAspTrpGlnLeuLeuAsnGlyProProThrM 474
1346 CCGCCTAGTGACAGGCTTCAGTTGG.....GTTCTGTCAG 1380
474 eLysThrSerProThrValTyProValAspAsnLeuProAlaArgPhe 490
1381 AGCAAGCAAGCTGGTTCTTCTTCCACTCAAGAACGAGAACGCTTAC 1430
491 ThrTyPhe 493
1431 CCGATCTTC 1439

seq_name: /SIDSL/gcdata/geneseq/geneseqn/NA2000.DAT:AAc48157
seq_documentation_block:
ID AAC48157 standard; DNA; 1646 bp.
AC AAC48157;
XX
XX 18-OCT-2000 (first entry)
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 56467.
XX
```

```
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX Arabidopsis thaliana.
XX EPI033405-A2.
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127452.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 19-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
XX 23-APR-1999; 99US-0130891.
XX 28-APR-1999; 99US-0131449.
XX 30-APR-1999; 99US-0132048.
XX 30-APR-1999; 99US-0132407.
XX 04-MAY-1999; 99US-0132484.
XX 05-MAY-1999; 99US-0132485.
XX 06-MAY-1999; 99US-0132486.
XX 06-MAY-1999; 99US-0132487.
XX 07-MAY-1999; 99US-0132863.
XX 11-MAY-1999; 99US-0134256.
XX 14-MAY-1999; 99US-0134219.
XX 14-MAY-1999; 99US-0134221.
XX 14-MAY-1999; 99US-0134370.
XX 18-MAY-1999; 99US-0134768.
XX 19-MAY-1999; 99US-0134941.
XX 20-MAY-1999; 99US-0135124.
XX 21-MAY-1999; 99US-0135353.
XX 24-MAY-1999; 99US-0135629.
XX 25-MAY-1999; 99US-0136021.
XX 27-MAY-1999; 99US-0136392.
XX 28-MAY-1999; 99US-0136782.
XX 01-JUN-1999; 99US-0137222.
XX 03-JUN-1999; 99US-0137528.
XX 04-JUN-1999; 99US-0137529.
XX 07-JUN-1999; 99US-0137702.
XX 08-JUN-1999; 99US-0138094.
XX 10-JUN-1999; 99US-0138540.
XX 10-JUN-1999; 99US-0138847.
XX 14-JUN-1999; 99US-0139119.
XX 16-JUN-1999; 99US-0139452.
XX 16-JUN-1999; 99US-0139453.
XX 17-JUN-1999; 99US-0139492.
XX 18-JUN-1999; 99US-0139454.
XX 18-JUN-1999; 99US-0139455.
XX 18-JUN-1999; 99US-0139456.
XX 18-JUN-1999; 99US-0139457.
XX 18-JUN-1999; 99US-0139458.
XX 18-JUN-1999; 99US-0139459.
XX 18-JUN-1999; 99US-0139460.
XX 18-JUN-1999; 99US-0139461.
XX 18-JUN-1999; 99US-0139462.
XX 18-JUN-1999; 99US-0139463.
XX 18-JUN-1999; 99US-0139750.
XX 18-JUN-1999; 99US-0139763.
XX 21-JUN-1999; 99US-0139817.
XX 22-JUN-1999; 99US-0139899.
XX 23-JUN-1999; 99US-0140353.
```

PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.

PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

alignment_scores: Quality: 450.00 Length: 503
 Ratio: 1.562 Gaps: 15
Percent Similarity: 57.256 Percent Identity: 26.640

alignment_block:
US-09-668-482-32 x AAC48157 ..

Align seg 1/1 to: AAC48157 from: 1 to: 1646

6 leuAlaIaseRAlaLeuCysThrPheValIeuProLeuLeuPheLe 22
|||||:||||| ::: |||::||| |||
83 CTCCTCTCTTCATCGCCGCCGCTTCTCTCTCTCTACAC..... 121
22 uAlaIaLeuLysIeuTrpAspLeuTrpCysValSerArgAspArg 39
122:.....CGCGTACACGTT 134
39 erCysAlaLeuProLeuProPogLyThrMetGlyPheProPheheGly 55
::: |||||||:|||||:|||||:|||||:|||||:|||||
135 ACCGTGCGATGGCTCGCTCCGCGAAGCCTTGCTCTCTCTGATPAGA 184
56 GluTrpLeuGlnMetVal.....LeuGlnArgArgGlySph 67
|||||:|||||:|||||:|||||:|||||:|||||
185 GAGACTTTTCACCTGATCGACCTTACAAACAGAGAACCTTGACCTTT 234

67 elenuimethysarqatgylsytyrllleuylphelleuythrlysthrleu 84
|||:::|||||:::|||||:::|||||:::|||||

235 catgcacagaaagatgaaccggcttagcggttccattcacacgatcttt 284
heglyatrprothrvalargvalmetglualaaspaasalargatgle 100
|||||:::|||||:::|||||:::|||||:::|||||

285 ttggtaaacccgcatgttttcacgttgacccgcaaacacacgggttgtt 334
101 leuleuglulinhsargleualseralhnstrpproalaservalar 117
|||:::|||||:::|||||:::|||||:::|||||

335 ctttcagaacgaaggaaacctttttgagtggtcttatcctgcgtccatttg 384
117 gthrleuleuglyalaaglcyseusleusasnleuhisasperserhist 134
:::|||||:::|||||:::|||||:::|||||

385 taaccttttgggaaaacactctctctcttcttatgaaagcttttgacata 434
134 ysglialrglysyalilemetginalapneserarglualaleuin 150
|||:::|||||:::|||||:::|||||:::|||||

435 aacgatagacactcttcacacatgacctttgctaattcttcataatattmaa 484
151 cystyryvalleual..tlelaiglueluaaserseycyleuglu 166
485 gaaccatctcatcttgatatgacccgggttagtcggtttaacttgatgcc 534
166 ntrpuesercysglguarglyleuaylleueualyrrproglualitysa 183
:::|||||:::|||||:::|||||:::|||||

535 ttggtctctt.....cgtgttctctctatggaagaccamaa 572
183 rglumeprhenargllelametarglleleuenglcysgaalnprogly 199
|||:::|||||:::|||||:::|||||:::|||||

573 acatpacctttagatgaacgggaagacagttagatagcctttgatcca... 619
200 prolaiglylglygluaspsluglnslneuleuvalgualaphleglu 216
|||||:::|||||:::|||||:::|||||

620ggggaatgagatgagatgattaaagaaagactattcttt 657
216 ukethrarglaslepserleuproilespuralpropheserlyl 233
|||:::|||||:::|||||:::|||||:::|||||

658 tctcatcgaagccttcttctctctctctctcctcctc....ttctccacca 704
233 eutytrarg..glyallysalaaargasnleuhishlaarlglelu 248
|||||:::|||||:::|||||:::|||||

705 cttaaaccgamaacccatccamacacggcgag.....aaagtggcg 742
249 gluasnileargalibysilleargrgleugnallathngluoproagl 265
|||:::|||||:::|||||:::|||||:::|||||

743 gagcgcttagcgtgtgtgtatgaaagacgaagacgaagacaagacg 792
265 ygly.....cyslyspaaleuenglneule.....gluhss 278
|||:::|||||:::|||||:::|||||:::|||||

793 aocggagagaaaaaagaatctcttgcggcggttgcttgccggcgatgag 842
278 etrtgluarglygluaeargleuaspmeglalaleuylsnlserser 294
:::|||||:::|||||:::|||||:::|||||

843 gattttccatgaadagatgttgacttc.....ttg 874
295 thrglueuleupheglylghiscluthrthralseralaalthse 311
:::|||||:::|||||:::|||||:::|||||

875 cggccttacttgcgcgcttatmaacacctccacagatcatcact 924
311 lleuethrtrylleuglyleutyrprohisvalleneuglnusvalarg 328
:::|||||:::|||||:::|||||:::|||||

925 ccgcgtcaaatcttcacccgaacactctttagccttctgtcacactcaag 974
328 lngluilleysserlysgleuleucyslysserasnlnaspalus 344
|||||:::|||||:::|||||:::|||||

975 aaag.....catgaaaaagatttagcgcamtgaadagatctcgtatagt 1018
345 leasmpetgluthreuglnleuynustyrillieglyasyvalieleu 361
|||||:::|||||:::|||||:::|||||

1019 cttgaaatgagatgattacaaagcatgcatgcattcacacatgtgtgtgtna 1066
361 scluthrleuaragleuasnpiovalproglglylpheargalatal 378

```

1069 TAGAGCGCTACGATCGGTACATCATCGCGCGGTTTTCAGACGTCAA 1118
      :::::::::::::::::::: :::: ||| ::::|::| ::::|
378 eulysThrhegluLeuasnGlyTyrGlnIleProlySGlyTPraSnVal 394
      :: ::::::::::::::::::::|::::::::::::|::::::::::::|
1119 TCACGGATTTGAGATCAAAAGGTTATATAATTCACAAAGGGTGAAGA 1168
      ::::::::::::::::::::|::::::::::::|::::::::::::|
395 IleTyrSerIleCysAspThrHisAspValAlaAspIlePheThrAsn 411
      ::::::::::::::::::::|::::::::::::|::::::::::::|
1169 TTCTCATCGTTTAGACGCGGTTCTATTAGACCCAAACCATTCACAAAGATGC 1218
      ::::::::::::::::::::|::::::::::::|::::::::::::|
411 sGluGluPheAsnProAspArgPhe.....IleValProHisP 424
      ::::::::::::::::::::|::::::::::::|::::::::::::|
1219 TCGCACTTTCACCCCTTCGAGATGGCAGACCACTCGGTACGACAGGCC 1268
      ::::::::::::::::::::|::::::::::::|::::::::::::|
424 roGluAspAlaSerArgPheSerPheIleProPheGlyGlyLeuArg 440
      ::::::::::::::::::::|::::::::::::|::::::::::::|
1269 CTCTATATGTG.....TTGCACACGTTTGGTGGAGGCGCAC 1306
      ::::::::::::::::::::|::::::::::::|::::::::::::|
441 SerCysValGlyGlySGluPheAlaValIleLeuLeuLysIlePheThr 457
      ::::::::::::::::::::|::::::::::::|::::::::::::|
1307 CTATGTCGGCGTTACAGCTGGCTAGGGTTCACACTCTCTGTTTCCTTCA 1356
      ::::::::::::::::::::|::::::::::::|::::::::::::|
457 IGLuLeuAlaArgHisCysAspTrrpGlnLeuLeuAsnGlyProProThrm 474
      ::::::::::::::::::::|::::::::::::|::::::::::::|
1357 CGCCCTAGAGACAGCGCTTCAGTGG.....GTTCCTGCAG 1391
      ::::::::::::::::::::|::::::::::::|::::::::::::|
474 eulysThrSerProThrValTyrProValAspAsnLeuProAlaArgPhe 490
      ::::::::::::::::::::|::::::::::::|::::::::::::|
1392 AGCAACACAAAGCTGTTTCTTCCAACTACAGACAGCAAGAAACGGTAC 1441
      ::::::::::::::::::::|::::::::::::|::::::::::::|
491 ThrTyrPhe 493
      |||
1442 CCGATCTTC 1450

seq_name: /SID1/gogdata/geneseq/geneseqn/NA2000.DAT:AC47986
seq_documentation_block:
ID AC47986 standard; DNA; 1649 BP.
XX AAC47986;
AC
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 55845.
XX
KW Hybridisation assay; genetic mapping; gene expression control,
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
CD 06-SEP-2000.
XX
FE 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127464.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 99US-0131449.

```


PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132488.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0138847.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 18-JUN-1999; 99US-0139817.
PR 21-JUN-1999; 99US-0139899.
PR 22-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 23-JUN-1999; 99US-0140695.
PR 24-JUN-1999; 99US-0140823.
PR 28-JUN-1999; 99US-0140991.
PR 29-JUN-1999; 99US-0141287.
PR 30-JUN-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 16-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 19-JUL-1999; 99US-0144336.
PR 20-JUL-1999; 99US-0144337.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.

PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149930.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151308.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159283.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.


```
474 etlysrhrserProthValtryProValaspasnlLeuProAlaArgPhe 490
      ::: ::: ::|||::: ::|
1394 AGCAAGACAACTGTTTCTTCCACTACAGACGACGAACGCGTAC 1443
      491 ThrTyrrPhe 493
      |||
1444 CCGATCTTC 1452

seq_name: /SIDSL/gcgdata/geneseq/geneseqn/NA2000.DAT: AAC35582
seq_documentation_block:
ID AAC35582 standard; DNA; 1667 BP.
XX AAC35582;
XX
XX 17-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 10705.
XX
XX Hybridisation assay; genetic mapping; gene expression control;
XX protein identification; signal transduction pathway;
XX metabolic pathway; promoter; termination sequence; ss.
XX Arabidopsis thaliana.
XX EPI033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 19-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
XX 23-APR-1999; 99US-0130891.
XX 28-APR-1999; 99US-0131449.
XX 30-APR-1999; 99US-0132048.
XX 30-APR-1999; 99US-0132407.
XX 04-MAY-1999; 99US-0132484.
XX 05-MAY-1999; 99US-0132485.
XX 06-MAY-1999; 99US-0132486.
XX 07-MAY-1999; 99US-0132863.
XX 11-MAY-1999; 99US-0134256.
XX 14-MAY-1999; 99US-0134218.
XX 14-MAY-1999; 99US-0134219.
XX 14-MAY-1999; 99US-0134221.
XX 14-MAY-1999; 99US-0134370.
XX 18-MAY-1999; 99US-0134768.
XX 19-MAY-1999; 99US-0134941.
XX 20-MAY-1999; 99US-0135124.
XX 21-MAY-1999; 99US-0135353.
XX 21-MAY-1999; 99US-0135629.
XX 24-MAY-1999; 99US-0136021.
XX 25-MAY-1999; 99US-0136392.
XX 27-MAY-1999; 99US-0136782.
XX 28-MAY-1999; 99US-0137223.
XX 01-JUN-1999; 99US-0137528.
XX 03-JUN-1999; 99US-0137502.
XX 04-JUN-1999; 99US-0137724.
XX 07-JUN-1999; 99US-0138094.
XX 08-JUN-1999; 99US-0138540.
XX 10-JUN-1999; 99US-0138847.
XX 10-JUN-1999; 99US-0138847.
```

```
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140693.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 28-JUL-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146389.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
```

```

PR 11-AUG-1999; 9905-0148319.
PR 12-AUG-1999; 9905-0148341.
PR 13-AUG-1999; 9905-0148565.
PR 13-AUG-1999; 9905-0148684.
PR 16-AUG-1999; 9905-0149368.
PR 17-AUG-1999; 9905-0149175.
PR 18-AUG-1999; 9905-0149426.
PR 20-AUG-1999; 9905-0149722.
PR 20-AUG-1999; 9905-0149723.
PR 20-AUG-1999; 9905-0149929.
PR 23-AUG-1999; 9905-0149902.
PR 23-AUG-1999; 9905-0149930.
PR 25-AUG-1999; 9905-0150566.
PR 26-AUG-1999; 9905-0150884.
PR 27-AUG-1999; 9905-0151065.
PR 27-AUG-1999; 9905-0151066.
PR 27-AUG-1999; 9905-0151080.
PR 30-AUG-1999; 9905-0151303.
PR 31-AUG-1999; 9905-0151438.
PR 01-SEP-1999; 9905-0151930.
PR 07-SEP-1999; 9905-0152363.
PR 10-SEP-1999; 9905-0153070.
PR 13-SEP-1999; 9905-0153758.
PR 15-SEP-1999; 9905-0154018.
PR 16-SEP-1999; 9905-0154039.
PR 20-SEP-1999; 9905-0154779.
PR 22-SEP-1999; 9905-0155139.
PR 23-SEP-1999; 9905-0155486.
PR 24-SEP-1999; 9905-0155659.
PR 28-SEP-1999; 9905-0156458.
PR 29-SEP-1999; 9905-0156596.
PR 04-OCT-1999; 9905-0157117.
PR 05-OCT-1999; 9905-0157865.
PR 06-OCT-1999; 9905-0157753.
PR 07-OCT-1999; 9905-0158029.
PR 08-OCT-1999; 9905-0158232.
PR 12-OCT-1999; 9905-0158569.
PR 13-OCT-1999; 9905-0159293.
PR 13-OCT-1999; 9905-0159294.
PR 13-OCT-1999; 9905-0159329.
PR 14-OCT-1999; 9905-0159329.
PR 14-OCT-1999; 9905-0159330.
PR 14-OCT-1999; 9905-0159331.
PR 14-OCT-1999; 9905-0159637.
PR 14-OCT-1999; 9905-0159638.
PR 18-OCT-1999; 9905-0159584.
PR 21-OCT-1999; 9905-0160741.
PR 21-OCT-1999; 9905-0160767.
PR 21-OCT-1999; 9905-0160768.
PR 21-OCT-1999; 9905-0160770.
PR 21-OCT-1999; 9905-0160814.
PR 21-OCT-1999; 9905-0160815.
PR 22-OCT-1999; 9905-0160980.
PR 22-OCT-1999; 9905-0160981.
PR 22-OCT-1999; 9905-0160981.
PR 25-OCT-1999; 9905-0161404.
PR 25-OCT-1999; 9905-0161405.
PR 25-OCT-1999; 9905-0161406.
PR 26-OCT-1999; 9905-0161359.
PR 26-OCT-1999; 9905-0161360.
PR 26-OCT-1999; 9905-0161361.
PR 28-OCT-1999; 9905-0161920.
PR 28-OCT-1999; 9905-0161922.
PR 28-OCT-1999; 9905-0161933.
PR 29-OCT-1999; 9905-0162142.

```

alignment_scores:

```

Quality: 423.50      Length: 486
Ratio: 1.523         Gaps: 16
Percent Similarity: 57.202      Percent Identity: 27.984

```

alignment_block:

US-09-668-482-32 x AAC35582 ..

Align seg 1/1 to: AAC35582 from: 1 to: 1667

```

1 MetGlyLeuProAlaLeuAlaSerAlaLeuCysThrPheValLeuPr 17
109 TTGGGCTCTCTGTGATCATGTT...TGTATATGACT..... 144
17 OleuLeuLeuPheLeuAlaLeuLeuLysLeuTrpAspLeuTYCysValS 34
145 .....GCTCTTCGCGATGCAACGATCGCATATT 175
34 erSerArgAspArgSerCysAlaLeuProLeuProGlyLysMetGly 50
176 CTAAAGAAAGT.....CTTCCTCTCTGGAACCTGGGC 207
51 PheProPhePheGlyLysThrLeuGlnMetValLeuGlnArgLysPh 67
208 TGGCCATATTTGTGAAACCACTGATTTCTTAACAAGCAAGCATTT 257
67 eleuGlnMetLysArgArgLysTYGlyPheLeuTYrThrHisLeuP 84
258 CATGAAAAACCAAGACTAAGATATGGAGTTTCTCAAGCTCCACTTC 307
84 heGlyArgProThrValArgValMetGlyAlaAspAsnValArgArgIle 100
308 TTGGCTGCCCAACATAGCTTCATATGAGCCAGACTTAACAGATACATT 357
101 LeuLeuGlyLysLysArgLeuValSerValHisTrpProAlaSerValAr 117
358 CTAATGAATGAATCGAAGAGACTTGTCGCGTTACCCGCAATCATGCT 407
117 gThrIleLeuGlyAlaGlyCysLeuSerAsnLeuHisAspSerHisSL 134
408 TGATATTTAGGAGACATCAACATAGCTGCGGTTCATGCGCCAGCCACC 457
134 ysgInArgLysLysValIleMetGlnAlaPheSerArgGluAlaLeuGln 150
458 GCGTAATAGAGAGCGCTGCTGCTTCTTAATAGCCCAACATGATGAA 507
151 CysTYrValLeuValIleAlaGlyLysValSerCysLeuGlyGlnTr 167
508 GACCATCTCTTG.....CCTAAGATTGATGATTCATGAAACTA 548
167 PleuSerCysGlyLysLysArgLysLeu.....ValYrProGly 181
549 TCTT...TGGCGTTGGGATGATCTTGACACGATGATATCCAAAGAA 595
181 alLysArgLeuMetPheArgIleAlaMetArgIleLeuLeuGlyCysGlu 197
596 CCAACATATGCAATTTTATCATCGTTGTACAATAATGCTGACACTTGG 645
198 ProGlyProAlaGlyLysGlyLysAspGlyGlnGlnLeuValGluAlaPh 214
646 AAAAAACCA.....GAGGTTGAAGAATATGACAGAGCTTTT 683
214 eglGlnMetThrArgAsnLeuPheSerLeuProIleAspValProPheS 231
684 CAAGCTTGTTGTGGAACTCTA...TCGGTCCGATGATATCCCGGAA 730
231 erGlyLeuTYrArgGlyValLysAlaArgAsnLeuIleHisAlaArgIle 247
731 CGAATTACCGCAGTGGAGTCCAGCAAGAAAC..... 762
248 GluGlnAsnIleArgAlaLysIleArgArgLeuGlnAlaThrGluProAs 264
763 .....AACATCGATAGTTATGACAGAACTGATGCAAGAAAGAAAGA 806
264 pGlyGly.....CysLysAspAlaLeuGlnLeuLeuIleGluHisSerT 279
807 GTCTGGAGAAACTTTCACAGACATGTTGGGTACTGATGAAGAGAGAG 856
279 rpgLuarGlyLysLysArgLeuAspMetGlnAlaLeuLysGlnSerSerT 295
:::||||| ||| ::| |||

```

```
857 ATAACGACTTGTGTAAACGATAAGAGATAGATCAAGTGTAAACG 906
296 GluLeuLeuPheGlyGlyHisGluThrAlaSerAlaIaThrSerLe 312
      ::::::::::::::::::::
907 ...ATCTTGTAATTCGGGTATAGAGACTGTCTCAACCTCCATGATGGC 953
312 uIleThrTyrLeuGlyLeuTyrProHisValLeuGlnLysValArgGlu 329
      ::::::::::::::::::::
954 TCTTAAGTATCTCCATGATCTCCAAAAGCTCTTGAAGAAGACTCAAGAAGAG 1003
329 LuIleLysSerLysGlyLeuLeuCysLysSerAsnGlnAspAsnLysLeu 345
      ::::::::::::::::::::
1004 AA.....CATTGGCTATAAGSAGAGAAAACGACCTGACGAACCGCTC 1047
346 AspMetGluThrLeuGlnGluGlnLeuLysTyrIleGlyCysValIleLysG 362
      ::::::::::::::::::::
1048 ACCTCGACGATATTAAATCATGAATTCACCTGACGCTGTGATCTTTGA 1097
362 uThrLeuArgLeuAsnProProValProGlyGlyPheArgValAlaLeuL 379
      ::::::::::::::::::::
1098 GACATCAAGATTTGGCAACGATTTGTTAATGTGTCTTAGGAAACTACTCTC 1147
379 ySThrPheGluLeuAsnGlyTyrGlnIleProLysGlyTrpAsnVal... 394
      ::::::::::::::::::::
1148 ACGACTTAGAACTCAACGCTTATTATCCCAAAAGTTGGAGAAATTAC 1197
395 IleTyrSerIleCysAspThrHisAspValAlaAspIlePheThrAsnLy 411
      ::::::::::::::::::::
1198 GTATACACAAAGAGAGATTAACTATGATCATCT...CTTATGAAGATCC 1244
411 sGluGluPheAsnProAspArgPheIleValProHisProGlnAspAlas 428
      ::::::::::::::::::::
1245 AATGATCTTTAAACCATGAGATGAGTGAAGAAAGCTTAGAA.....T 1288
428 eArgPheSerPheIleProPheGlyGlyLeuArgSerCysValGly 444
      ::::::::::::::::::::
1289 CAAGAGGCTATTCTTACTCTTTGGAGGTGAGTTAGCCTTGCCCTGGA 1338
445 LysGluPheAlaLysIleLeuLeuLysIlePheThrValGluLeuAlaAr 461
      ::::::::::::::::::::
1339 AAGGAACTAGGAATCTCGAAGTCTCAAGCTTCCCTCACTACTTGTGTTAC 1388
461 gHisCysAspTrpGlnLeuLeuAsnGlyProProThrMetLysThrSerP 478
      ::::::::::::::::::::
1389 AAAATATATGATGGGAA...GAGAAAGGAGAAACAATTAATGTCCTTTC 1435
478 roThrVal 480
      ::::::::::
1436 CAAGAGTT 1443
```


OM of: US-09-668-482-32 to: Issued_Patents_NA:* out_format : pfs
Date: Nov 6, 2001 3:39 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame-p2n.model -DEV=xlp
-O=/cgn2_1/USPTO/US09668482/runat_05112001_133353_10897/app-query.fasta.1.1680
-B=Issued_Patents_NA -QFMT=fasta -SUFFIX=rn1 -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELTEXT=7.000 -START=1 -MATRIX=blonsum62
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=200000000
-USER=US09668482@cgn1_1.104 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPHY
-WAIT -THREADS=1

Search information block:

Query: US-09-668-482-32
Query length: 497
Database: Issued_Patents_NA:*
Database sequences: 351203
Database length: 113238999
Search time (sec): 149.780000

score_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation	..	0.0
/cgn2_6/prodata/2/ina/5b_COMB.seq:US-08-882-164D-31				2612.00	5657.71		1725
/cgn2_6/prodata/2/ina/5b_COMB.seq:US-08-724-466B-5				2437.00	5277.52		1494
/cgn2_6/prodata/2/ina/5b_COMB.seq:US-08-882-164D-5				2437.00	5277.52		1494
/cgn2_6/prodata/2/ina/5b_COMB.seq:US-08-882-164D-38				1853.00	3988.67		1464
/cgn2_6/prodata/2/ina/5b_COMB.seq:US-08-724-466B-3				1734.00	3739.68		1850
/cgn2_6/prodata/2/ina/5b_COMB.seq:US-08-882-164D-3				1734.00	3739.68		1850
/cgn2_6/prodata/2/ina/5b_COMB.seq:US-08-882-164D-36				1251.00	2680.12		2677
/cgn2_6/prodata/2/ina/5b_COMB.seq:US-08-724-466B-11				604.00	1294.56		351
/cgn2_6/prodata/2/ina/5b_COMB.seq:US-08-882-164D-11				450.00	937.97		1608
/cgn2_6/prodata/2/ina/5b_COMB.seq:US-08-991-677-3				358.50	736.07		1883
/cgn2_6/prodata/2/ina/5b_COMB.seq:US-09-091-432-1				337.00	689.45		1838
/cgn2_6/prodata/2/ina/5b_COMB.seq:US-09-103-840A-2				329.00	568.01		4403765
/cgn2_6/prodata/2/ina/5b_COMB.seq:US-09-103-840A-1				329.00	567.99		4411529
/cgn2_6/prodata/2/ina/5b_COMB.seq:US-09-103-840A-2				317.00	541.81		4403765
/cgn2_6/prodata/2/ina/5b_COMB.seq:US-09-103-840A-1				317.00	541.79		4411529
/cgn2_6/prodata/2/ina/5b_COMB.seq:US-08-948-564-15				300.00	608.71		1831
/cgn2_6/prodata/2/ina/5b_COMB.seq:US-08-948-564-3				285.00	577.02		1691
/cgn2_6/prodata/2/ina/5b_COMB.seq:US-09-292-768-67				270.50	544.85		1757
/cgn2_6/prodata/2/ina/5b_COMB.seq:US-09-292-768-8				261.50	525.91		1665
/cgn2_6/prodata/2/ina/5b_COMB.seq:US-09-292-768-3				261.50	525.91		1665
/cgn2_6/prodata/2/ina/5b_COMB.seq:US-09-292-768-63				261.50	525.91		1665
/cgn2_6/prodata/2/ina/5b_COMB.seq:US-08-948-564-5				255.50	511.49		1644
/cgn2_6/prodata/2/ina/5b_COMB.seq:US-08-948-564-1				255.50	511.49		1644
/cgn2_6/prodata/2/ina/5b_COMB.seq:US-09-091-432-3				255.50	497.71		1516
/cgn2_6/prodata/2/ina/5b_COMB.seq:US-08-882-164D-37				250.50	501.89		1663
/cgn2_6/prodata/2/ina/5b_COMB.seq:US-08-882-164D-3				250.50	501.89		1666
/cgn2_6/prodata/2/ina/5b_COMB.seq:US-08-285-309-24				250.50	501.89		1666
/cgn2_6/prodata/2/ina/5b_COMB.seq:US-08-502-046-24				250.50	501.89		1666
/cgn2_6/prodata/2/ina/5b_COMB.seq:US-08-606-505B-63				249.00	495.06		2174
/cgn2_6/prodata/2/ina/5b_COMB.seq:US-09-091-432-3				249.00	495.06		2174
/cgn2_6/prodata/2/ina/5b_COMB.seq:US-09-292-768-63				246.50	492.41		1762
/cgn2_6/prodata/2/ina/5b_COMB.seq:US-09-292-768-63				246.50	492.41		1762
/cgn2_6/prodata/2/ina/5b_COMB.seq:US-08-881-784-5				241.50	481.49		1762
/cgn2_6/prodata/2/ina/5b_COMB.seq:US-09-292-768-1				241.50	481.49		1762
/cgn2_6/prodata/2/ina/5b_COMB.seq:US-09-292-768-65				241.00	480.40		1761
/cgn2_6/prodata/2/ina/5b_COMB.seq:US-09-033-055A-3				241.00	480.40		1762
/cgn2_6/prodata/2/ina/5b_COMB.seq:US-09-172-339-5				241.00	480.40		1762

/cgn2_6/prodata/2/ina/5b_COMB.seq:US-08-948-564-17 + 238.50 475.39 7.9e-19 1704
/cgn2_6/prodata/2/ina/5b_COMB.seq:US-08-194-981E-3 + 237.50 474.12 9.3e-19 1591
/cgn2_6/prodata/2/ina/5b_COMB.seq:US-08-948-564-9 + 237.50 472.56 1.1e-18 1788
/cgn2_6/prodata/2/ina/5b_COMB.seq:US-08-201-116-10 + 237.50 472.09 1.2e-18 1852
/cgn2_6/prodata/2/ina/5b_COMB.seq:US-08-238-821B-10 + 237.50 472.09 1.2e-18 1852

seq_name: /cgn2_6/prodata/2/ina/5b_COMB.seq:US-08-882-164D-31

seq_documentation_block:

Sequence 31, Application US/08882164D
Patent No. 6306624
GENERAL INFORMATION:
APPLICANT: Petkovich, P. Martin, White, Jay A.,
TITLE OF INVENTION: Retinoid Metabolizing Protein
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Blake, Cassels & Graydon
STREET: Box 25, Commerce Court West
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5L 1A9
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 Mb storage
COMPUTER: COMPAQ, IBM PC compatible
OPERATING SYSTEM: MS-DOS 5.1
SOFTWARE: WORD PERFECT
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/882,164D
FILING DATE: June 25, 1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/667,546
FILING DATE: June 21, 1996
APPLICATION NUMBER: 08/724,466
FILING DATE: October 1, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Hunt, John C.
REGISTRATION NUMBER: 36,424
REFERENCE/DOCKET NUMBER: 50767/00010
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 863-4344
TELEFAX: (416) 863-2653
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 1725 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-882-164D-31

alignment_scores:

Quality: 2612.00 Length: 497
Ratio: 5.256 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-668-482-32 x US-08-882-164D-31 ..

Align seg 1/1 to: US-08-882-164D-31 from: 1 to: 1725

1 MetGlyLeuProAlaLeuAlaSerAlaLeuCysThrPheValLeuPr 17
|||||
25 ATGGGCTCCCGGCTGCTGGCCACTGGCTTGCACCTTCGTGGCC 74
17 oleuLeuPheLeuAlaAlaLeuLysLeuTPASPLeuTyrcysValS 34
|||||
75 GCGTGGCTTCTTCTGGCGGCTCAAGCTGTGGACCTGTATGTGTGA 124
34 eSerArgAspArgSerCysAlaLeuProLeuProGlyThrMetGly 50
|||||
125 GCAGCGCGATCGACATGGCGCTCCCTTGGCCCGGCTACATGGGC 174

```

51 PheProPheGlyGluThrLeuGlnMetValLeuGlnArgArgLysPhe 67
175 TTCACATCTTTGGGAAACATTCAGATGGTGTTCACGGAGAGAGATT 224
67 LeuGlnMetLysArgArgLysTyrGlyPheIleTyrLysThrHisLeuP 84
225 TCTGAGATGAAGCGCAGAAATACGGCTTCATCAAGACGATCTGT 274
84 heGlyArgProThrValArgValMetGlyAlaAspAsnValArgArgIle 100
275 TTGGGGCCCAACGGTGGGGGTGATGGCGGATTAATGCGGCGCANTC 324
101 LeuLeuGlyGlnHisArgLeuValSerValHisTyrProAlaSerValAr 117
325 TTGCTGGAGAGACACGGTGGTGGTGGTGCACCTGCCCGGCTGGTGG 374
117 GThrIleLeuGlyAlaGlyCysLeuSerAsnLeuHisAspSerHisSL 134
375 CACCATCTGGGGCGCTGGCTGCTCCACCTGCACGATTCCTCGACAC 424
134 YSGlnArgLysLysValIleMetGlnAlaPheSerArgGluAlaLeuGln 150
425 AGCAGCGAAAGAGGTGATTATGACGGCTTCAGCGCGAGGCACTCCAG 474
151 CysTyrValLeuValIleAlaGlnGluValSerSerCysLeuGlnIle 167
475 TGCTACGCTGCTGATGCTGAGAGAGTCAAGCATGTTGCTGGACACTG 524
167 PLeuSerCysGlyGluArgGlyLeuLeuValTyrProGluValLysArgL 184
525 GCTAAGCTGGCGGCGGCGGCTCTGCTTACCCGAGGGAAGCCGCC 574
184 eumetPheArgIleAlaMetArgIleLeuLeuGlyCysGluProGlyPro 200
575 TCATGTTCCGATCGCATGGCGATCCCTGGGTGGGTGGAGCCGGGTCCA 624
201 AlaGlyGlyGlyGluAspGluGlnGlnLeuValGluAlaPheGluGlu 217
625 GCGGCGGCGGAGAGAGACAGCAACAGCTCGTGAGGCTTCGAGAGAT 674
217 tThrArgAsnLeuPheSerLeuProIleAspValProPheSerGlyLeu 234
675 GACCCGCAATCTTCTCTCTCCATTCAGCTGCTTTCAGCGGCTGT 724
234 YrArgGlyValLysAlaArgAsnLeuIleHisAlaArgIleGluGluAsn 250
725 ACCGGGGGTGAAGCGCGGAACCTTATACAGCGGCATCGAGAGAAC 774
251 IleArgAlaLysIleArgArgLeuGlnAlaThrGluProAspGlyGly 267
775 ATTGCGCCCAAGATCCGCGCTTCAGGCTACAGAGCGGATGGGGTGG 824
267 sLysAspAlaLeuGlnLeuLeuIleGluHisSerTyrGlnArgGlyGlu 284
825 CAGAGAGCGGCTGCACTCTGATTCAGCACTGCTGGAGAGGGAGAGA 874
284 rGLeuAspMetGlnAlaLeuLysGlnSerSerThrGluLeuLeuPheGly 300
875 GGGTGGATATGACAGCACTAAACAATCGTCAAGAGACTCTCTTGGT 924
301 GlyHisGluThrThrAlaSerAlaIleThrSerLeuIleThrTyrLeu 317
925 GGTGATGAACACTACAGCCAGTGGTGCACATCACTGATCACTACAG 974
317 YLeuTyrProHisValLeuGlnLysValArgGluGluIleLysSerLys 334
975 ACTGTACCCACATCTCTCCAGAAAGTTTCAGAGAGATTAAGAGCAAG 1024
334 YLeuLeuCysLysSerAsnGlnAspAsnLysLeuAspMetGluThrLeu 350
1025 GCTTACTTTCAGAGCAATCAAGACAAACAAGTTAGCATGGAACCTTGG 1074

```

```

351 GlnGlnLeuLysTyrIleGlyCysValIleLysGluThrLeuArgLys 367
1075 GAACAGCTTAATACATTTGGGTGTGTCATTTAAGAGACCTCGATGAA 1124
367 nProProValProGlyGlyPheArgValAlaLeuLysThrPheGluLeu 384
1125 TCCTCCGGTTCAGAGAGGGTTCCGGTTGCTCTGAAGACTTTTGAGCTGA 1174
384 sngLysTyrGlnIleProLysGlyTyrPasnValIleTyrSerIleCys 400
1175 ATGGATACCAAGATCCCAAGGGCTGGAATGTTATTACAGATCTGTGAC 1224
401 ThrHisAspValAlaAspIlePheThrAsnLysGluGluPheAsnProAs 417
1225 ACCCAGCATGTGGCAGATATCTTCACTAACAAGAGGAATTAATCCCA 1274
417 PaRgPheIleValProHisProGluAspAlaSerArgPheSerPheIle 434
1275 CCGCTTATAGTCCCTCATCCAGAGGATGCTTCCGGTTCACCTTCATTC 1324
434 rOPheGlyGlyLeuArgSerCysValGlyLysGluPheAlaLysIle 450
1325 CATTTGAGAGAGCCCTTGGAGCTGTGTAGGCAGAAAGCTTGCAGAAAT 1374
451 LeuLeuLysIlePheThrValGluLeuValaArgHisCysAspTyrGln 467
1375 CTCTTAAGATATTACAGTGGAGCTGGCTAGGCACCTGTGATGGCAGCT 1424
467 uLeuAsnGlyProProThrMetLysThrSerProThrValTyrProVal 484
1425 TCTAATAGACTCTTACAAATGAAAGCAAGCCCACTGTGACCTGTGG 1474
484 sPasnLeuProAlaArgPheThrTyrPheGlnGlyAspIle 497
1475 ACAATCTCCCTGCAGATTCACCTACTTCCAGGAGATATC 1515

seq_name: /cgn2.6/ptodata/2/lna/6A_COMB.seq:US-08-724-466B-5
seq_documentation_block:
; Sequence 5, Application US/08724466B
; Patient No. 6063606
; GENERAL INFORMATION:
; APPLICANT: Petkovich, P. Martin, White, Jay A.,
; APPLICANT: Beckovitch, Barbara R., Jones, Glenville
; TITLE OF INVENTION: Retinoid Metabolizing Protein
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Blake, Cassels & Graydon
; STREET: Box 25, Commerce Court West
; CITY: Toronto
; ZIP: M5L 1A9
; COUNTRY: Canada
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 Mb storage
; OPERATING SYSTEM: MS-DOS 5.1
; SOFTWARE: WORD PERFECT
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/724,466B
; FILING DATE: October 1, 1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/667,546
; FILING DATE: June 21, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Hunt, John C.
; REGISTRATION NUMBER: 36,424
; REFERENCE/DOCKET NUMBER: 50767/00004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 863-4344
; TELEFAX: (416) 863-2653
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1494 base pairs

```



```

;
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-724-466B-5

```

```

alignment_scores:
    Quality: 2437.00      Length: 497
    Ratio: 5.035          Gaps: 0
    Percent Similarity: 97.384    Percent Identity: 93.360

```

```

alignment_block:
US-09-668-482-32 x US-08-724-466B-5 ..

```

```

Align seg 1/1 to: US-08-724-466B-5 from: 1 to: 1494

```

```

1 MetGluLeuProAlaLeuAlaSerAlaLeuCysThrPheValLeuPr 17
1 ATGGGGCTCCCGCGCTGCTGGCCAGTGCCTTCGACCTTCGCTGCC 50
17 OleuLeuLeuPheLeuAlaAlaLeuLysLeuTrpAspLeuTrpCysValS 34
1 GCTGCTGCTCTTCTGCTGGCTGACATCAAGCTCTGGACCTGACTGCGTGA 100
34 erSerArgAspArgSerCysAlaAlaLeuProLeuProProGlyThrMetGly 50
101 GCGGCGCGGACCGCAGTGTGCTCCCTCCATTGCCCCCGGAGACTATGGC 150
51 PheProPhePheGlyGluThrLeuGluMetValLeuGluInArgArgLysPhe 67
151 TTCCTCTTCTTGGGGAAACCTTCGACATGCTACTGCAGCGAGAGAAAGTT 200
67 eLeuGluMetLysArgArgLysTrpGlyPheIleTrpLysThrIleLeuP 84
201 CCTGAGATGGAAGCGCAGGAATACGGCTTCATCTACAGACCATCTGT 250
84 heGlyArgProThrValArgValMetGlyAlaAspAsnValArgArgIle 100
251 TGGGGCGGCCACCGTACGGGTATGGCGCGGACATGTGCGGCGCATC 300
101 LeuLeuGlyGluHisArgLeuValSerValHisTrpProAlaSerValAr 117
301 TTGCTCGGAGAGACCGCGCTGTGTGCGGTCCACTGGCCAGCGTGTGCG 350
117 gThrIleLeuGlyAlaGlyCysLeuSerAsnLeuHisAspSerSerHisL 134
351 CACCAFTCTGGGATCTGCTGCTCTTACCTCAGACATCTCTCGCACCA 400
134 ySGLnArgLysLysValIleMetGlnAlaPheSerArgGluAlaLeuGln 150
401 AGCAGCGCAGAGAGGTATATTCGGGCTTCAGCGCGGAGGACTCGAA 450
151 CysTrpValLeuValIleAlaGluGluValSerSerCysLeuGluGlnTr 167
451 TGCTACGTGCGCGGTGATCACCAGAGAGATGGGACAGCCTCGAGACAGTG 500
167 pleuSerCysGlyGluArgGlyLeuLeuValTrpProGluValLysArgL 184
501 GCTGAGCTGCGCGCAGCGCGGCTCTCTGCTACCCGAGGTGAAGCGCC 550
184 eumPhePheArgIleAlaMetArgIleLeuLeuGlyCysGluProGlyPro 200
551 TCATGTTCGGAATCCCATGCGCATCTTACTGCTGGCGGAACCCCAACTG 600
201 AlaGlyGlyGlyLysAspGluGlnGlnLeuValGluAlaPheGluGluMe 217
601 GCGGGGAGAGGGAGACTCCGACAGCAGCAGCTTGAGAGGCTTCGAGAAAT 650
217 tThrArgAsnLeuPheSerLeuProIleAspValProPheSerGlyLeuTr 234
651 GAACCCCAATCTCTTCTGCTGCTCCCATGACGTGACCTTCACGCGGCTGT 700
234 yArgGlyValLysAlaArgAsnLeuIleHisAlaArgIleGluGluAsn 250

```

```

|||||.....
701 ACCGGGCGATGAGCGCGGAGAACCTCATTCACGGCGCATGAGACAGAAC 750
251 IleArgAlaLysIleArgArgLeuGlnAlaThrGluProAspGlyGlyC 267
|||||.....
751 ATTCGGCGCAAGATCTGCGGCTCGGCGCATCCGAGCGGCGCAAGTTGG 800
267 sLysAspAlaLeuGlnLeuIleGluHisSerTrpGluArgGlyGluA 284
801 CAAGAGCGCGTGCAGCTGTGATCGACACTCTGGGAGAGGGAGAGAC 850
|||||.....
851 GGCTGACATGACAGGACATAAGCATCTTCAACCAACTCCTTTGGA 900
301 GlyHisGluThrThrAlaSerAlaIleThrSerLeuIleThrTyrLeuG 317
901 GGACAGAAACCAAGCGCCAGTGCAGCCACATCTGTATCATCTTACCTGG 950
317 yLeuTrpProHisValLeuGluLysValArgGluGluIleLysSerLysG 334
951 GCTTACCCACATGTTCTCCAGAAAGTGGAGAGAGCTGAAGAGTAAG 1000
334 yLeuLeuCysLysSerAsnGlnAspAsnLysLeuAspMetGluThrLeu 350
1001 GTTACTTTGCAAGACCAATCAAGACACAGATTGGACATGGAAATTGG 1050
351 GluGlnLeuLysTrpIleGlyCysValIleLysGluThrLeuArgLeu 367
1051 GAACCACTTAATACATCGGGTGTGTTATTAAAGAGACCTTCGACTGAA 1100
367 nProProValProGlyGlyPheArgValAlaLeuLysThrPheGluLeuA 384
1101 TCCCCAGTTCACAGAGAGGTTTGGGGTGTCTGAAAGACTTTGAATTAA 1150
384 snGlyTrpGlnIleProLysGlyTrpAsnValIleTrpSerIleCysAsp 400
1151 ATGATACCAATTCACCAAGGCTGGAATGTATCTACATCTGTGAT 1200
401 ThrHisAspValAlaAspIlePheThrAsnLysGluLupheAsnProAs 417
1201 ACCTATATGTGGCAGAGATCTTCACCAACAGAGAAATTTAATCTGA 1250
417 ParGpheIleValProHisProGluAspAlaSerArgPheSerPheIleP 434
1251 CCGATTGAGTGTCTTCACCCAGAGATGCATCCAGGTTCAGCTTCATTC 1300
434 ropheGlyGlyGlyLeuArgSerCysValGlyLysGluPheAlaLysIle 450
1301 CATTTGAGAGGAGCGCTTGAAGAGCTGTGTAGCAAAAGATTGCAAAATT 1350
451 LeuLeuLysIlePheThrValGluLeuAlaArgHisCysAspTrpGlnLe 467
1351 CTCTGCAAAATATTACAGTGGAGACTGGCCAGGATTTGATCGGAGCT 1400
467 uLeuAsnGlyProProThrMetLysThrSerProThrValTrpProValA 484
1401 TCTAATGAGACCTCTTACATGAAGAACAGTCCACCGTGTATCTGTGG 1450
484 spAsnLeuProAlaArgPheThrTrpPheGlnGlyAspIle 497
1451 ACAATCTCCCTGCAAGATTCCACCATTTCCATGGGGAATTC 1491
seq_name: /cgn2_6/prodata/2/lna/6B_COMB.seq:US-08-882-164D-5
seq_documentation_block:
; Sequence 5, Application US/08882164D
; Patent No. 6306624
; GENERAL INFORMATION:
; APPLICANT: Petkovich, P. Martin, White, Jay A.,
; APPLICANT: Beckett, Barbara R., Jones, Glenville
; TITLE OR INVENTION: Retinoid Metabolizing Protein
; NUMBER OF SEQUENCES: 43

```

```

CORRESPONDENCE ADDRESS:
ADDRESS: Blake, Cassels & Graydon
STREET: Box 25, Commerce Court West
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5L 1A9
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 Mb storage
COMPUTER: COMPAQ, IBM PC compatible
OPERATING SYSTEM: MS-DOS 5.1
SOFTWARE: WORD PERFECT
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/882,164D
FILING DATE: June 25, 1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/667,546
FILING DATE: June 21, 1996
APPLICATION NUMBER: 08/724,466
FILING DATE: October 1, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Hunt, John C.
REGISTRATION NUMBER: 36,424
REFERENCE/DOCKET NUMBER: 50767/00010
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 863-4344
TELEFAX: (416) 863-2653
INFORMATION FOR SEO ID NO: 5
SEQUENCE CHARACTERISTICS:
LENGTH: 1494 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-882-164D-5

alignment_scores:
Quality: 2437.00      Length: 497
Ratio: 5.035          Gaps: 0
Percent Similarity: 97.384      Percent Identity: 93.360

alignment_block:
US-09-668-482-32 x US-08-882-164D-5 ..
Align seg 1/1 to: US-08-882-164D-5 from: 1 to: 1494

1 MetGlyLeuProAlaLeuAlaSerAlaLeuCyThrPheValLeuPr 17
|||||
1 ATGGGGCTCCGGCGGTGGCGGCGAGTGGCTGTGCACCTTGGTGGTGC 50
17 OleuLeuPheLeuAlaLeuLysLeuTyrPaspLeuTyrCysValS 34
|||||
51 GCTGCTGCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 100
34 erSerArGspArGserCysAlaLeuProLeuProProGlyThrMetGly 50
|||||
101 GCGGCGCGCACCGCAGTGTGCTCCATTTGCCCCCGGAGCTATGGGC 150
51 PheProPheheGlyLthrLeuGlnMetValLeuGlnArGArgLysP 67
|||||
151 TTCCCTCTCTTTGGGAAACCTTGAGATGTACTGAGCGGAGGAAGTT 200
67 eleuGlnMetLysArGArgLysTyrGlyPheLeuTyrLysThrHisLeu 84
|||||
201 CCTCGAGATGAAGCGAGAAATACGCTTCACTACAGACCATCTGT 250
84 heGlyArGProThrValArGValMetGlyAlaAspAsnValArGArgL 100
|||||
251 TCGGGGGGCGCACCGCTACGGGTATGGCGGAGCAATGTGCGGCGATC 300
101 LeuLeuGlyGluHisArgLeuValSerValHisTrpProAlaSerValAr 117
|||||
301 TTGCTCGAGAGACGCGCTGTGTGCTGCTCACTGCGCAGCGTGGTGGG 350
```

```

117 gThrIleLeuGlyAlaGlyCysLeuSerAsnLeuHisAspSerSerHisL 134
|||||
351 CACCATTTCTGGGATCTGGCTGCTCTTAACCTGCACGACTCTCTGCACA 400
134 ySgInArGlysLysValIleMetGlnAlaPheSerArgGluAlaLeuGln 150
|||||
401 AGCAGCGCAGAAAGGTGATTTATGCGGGCTTCAGCCGAGGAGCTCGCAA 450
151 CysTyrValLeuValIleAlaGluGluValSerSerCysLeuGlnGluTr 167
|||||
451 TGCTACGTCGCGGTATATCACCAGAGAAATGGCGACACCTTGAGACAGT 500
167 PleuSerCysGlyGluArGlyLeuLeuValTyrProGluValLysArgL 184
|||||
501 GCTGAGTGTGCGGCGAGCGGCTCTCTGTTACCCGAGGAGGAAGCGCC 550
184 eUmetPheArgIleAlaMetArgIleLeuLeuGlyCysGluProGlyPro 200
|||||
551 TCATGTTCCGAATGCGCATGCGCATCTTACTGTGGCTGCGAAACCCCACTG 600
201 AlaGlyGlyGlyGluAspGluGlnGlnLeuValGluAlaPheGluGlu 217
|||||
601 GCGGCGCAGCGGAGACTCCGACAGACAGCTTGTGAGGCGCTTGAGAAAT 650
217 tThrArGAsnLeuPheSerLeuProIleAspValProPheSerGlyLeuT 234
|||||
651 GACCGCAATCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 700
234 YrArGlyValLysAlaArGAsnLeuIleHisAlaArgIleGluGluAsn 250
|||||
701 ACCGGGATGAAAGCGCGGAGACTCATTTACGCGCGCATGACAGACAGAC 750
251 IleArGAlaLysIleArGArgLeuGlnAlaThrGluProAspGlyGlyCy 267
|||||
751 ATTGCGCGCAAGATCTGCGGCTGCGGCGCATCGAGCGGCGCCAGGCTGT 800
267 sLysAspAlaLeuGlnLeuLeuIleGluHisSerTrpGluArGlyGlu 284
|||||
801 CAAGACCGCGCTGCAGCTGTTGATGAGACACTGCTGGAGAGGAGAGAC 850
284 rGleuAspMetGlnAlaLeuLysGlnSerSerThrGluLeuLeuPheGly 300
|||||
851 GGCTGACATGACGACGACCTAAAGCAATCTTCAACCGAATCTCTTTGGA 900
301 GlyHisGlnThrThrAlaSerAlaIleThrSerLeuIleThrTyrLeuG 317
|||||
901 GGACACGAAACACAGCGCAGTGCACACATCTCTGTATCACTTACCTGGG 950
317 yLeuTyrProHisValLeuGlnLysValArGluGluIleLysSerLysG 334
|||||
951 GCTTACCCACATGTTTCTCCAGAAAGTGCAGAAAGCTGTAAGAGTAAG 1000
334 lYleuLeuCyLysSerAsnGlnAspAsnLysLeuAspMetGluThrLeu 350
|||||
1001 GTTTACTTTCCAAGAGCAATCAAGACAAAGTTGGACATGAGAAATTTTG 1050
351 GluGlnLeuLysTyrIleGlyCysValIleLysGlnThrLeuArgLeuAs 367
|||||
1051 GAACCACTTAATATCATCGGCTGTGTTATTAAAGAGACCTTGCAGCTGAA 1100
367 nProProValProGlyGlyPheArGValAlaLeuLysThrPheGluLeu 384
|||||
1101 TCCCGCAGTTCAGAGAGGTTTGGGGTGTCTGGAAGACTTTGAATTA 1150
384 snGlyTyrGlnIleProLysGlyTyrPasnValIleTyrSerIleCysAsp 400
|||||
1151 ATGGATACCAAGATTTCCCAAGGCTGGAATGTATCTACACTATCTGTGAT 1200
401 ThrHisAspValAlaAspIlePheThrAsnLysGluGluPheAsnProAs 417
|||||
1201 ACTCATGATGTGSCAGAGATCTTACCAACAGAGAAATTTATATCTGA 1250
```

```

417 PAtGPheIleValProHISProGluAspAlaSerArgPheSerPheIleP 434
||||| :|||
1251 CCGATTGAGTCTCTCCACCCAGAGATGCATCCAGGTTCCGTTCAATTC 1300
434 roPheGlyGlyGlyLeuArgSerCysValGlyGluPheAlaLysIle 450
|||||
1301 CATTTGAGGAGGCGCTTAGAGAGCTGTAGGCAAAATTTGCAAAATTT 1350
451 leuLeuLysIlePheThrValGluLeuAlaArgHisCysAspTrpGlnLe 467
|||||
1351 CTCTCAAAATATTACAGTGGAGCTGGCCAGGCAATGTGACTGGCAGCT 1400
467 uLeuAsnGlyProProThrMetLysThrSerProThrValTrpProVala 484
|||||
1401 TCTAAATGAGACTCTCTCAATGAAGAAACAGTCCACCGCTATCTCTGTGG 1450
484 sPaSnLeuProAlaArgPheThrTrpPheGlnGlyAspIle 497
|||||
1451 ACAATCTCCCTGCAGAAATTCACCCATTTCCATGGGAAATTC 1491

```

seq_name: /cgn2_6/ptodata/2/lna/6B_COMB.seq:US-08-882-164D-38

```

seq_documentation_block:
; Sequence 38, Application US/08882164D
; Patent No. 6306624
; GENERAL INFORMATION:
; APPLICANT: Petkovich, P. Martin, White, Jay A.,
; APPLICANT: Beckett, Barbara R., Jones, Glenville
; TITLE OF INVENTION: Retinoid Metabolizing Protein
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Blake, Cassels & Graydon
; STREET: Box 25, Commerce Court West
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5L 1A9
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 Mb storage
; COMPUTER: COMPAQ, IBM PC compatible
; OPERATING SYSTEM: MS-DOS 5.1
; SOFTWARE: WORD PERFECT
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/882,164D
; FILING DATE: June 25, 1997
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 08/667,546
; FILING DATE: June 21, 1996
; APPLICATION NUMBER: 08/724,466
; FILING DATE: October 1, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Hunt, John C.
; REGISTRATION NUMBER: 36,424
; REFERENCE/DOCKET NUMBER: 50767/00010
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 863-4344
; TELEFAX: (416) 863-2653
; INFORMATION FOR SEQ ID NO: 38:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4164 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-882-164D-38

```

alignment_scores:

Quality: 1853.00 Length: 1046
Ratio: 4.240 Gaps: 10
Percent Similarity: 41.778 Percent Identity: 41.396

alignment_block:

US-09-668-482-32 x US-08-882-164D-38

Align seg 1/1 to: US-08-882-164D-38 from: 1 to: 4164

```

40 CysAlaLeuProLeuProProGlyThrMetGlyPheProPhePheGlyI 56
||||| :|||
73 TCGGCTCTCTCTCCACCTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCT 99
56 uThrLeuGlnMetValLeuGlnArgArgLysPheLeuGlnMetLysArg 73
|||||
100 .....CAGCGGAGGAAGTTTCTGCAGATGAAGCGCA 130
73 rGlySerGlyPheIleTrpLysThrHisLeuPheGlyArgProThrVal 89
|||||
131 GGAATFACGGCTTCAATCAAGAGCATCTGTGGGGGCCACGGGTG 180
90 ArgValMetGlyAlaAspAsnValArgArgIleLeuLeuGlyLysHis 106
|||||
181 CGGGGATGGGGCGGATATGTGGGGGCACTTGTGGGGAGACCGG 230
106 gLeuValSerValHisTrpProAlaSerValArgThrIleLeuGlyAla 123
|||||
231 GTGGTGTGGGTGCACGTGGCCCGCTGGTGGCCACCATCTGGGCGCTG 280
123 lYcysLeuSerAsnLeuHisAspSerSerHisLysGlnArgLys..Lys 138
|||||
281 GGTGCTCTCCAACTGCACGATTCCTCGCACAGCAGGAAAGAGGTG 330
138 ..... 138
331 AGGGTAGCTGGCAACTCTTGGCTGGCAGGAGACCTCATCTATGGCT 380
138 ..... 138
381 TGGTTCAGCAAAATAGAAATGGGGCGGAGGCTAGTCTATGTGTGGG 430
138 ..... 138
431 GACCAAGACCTCTCTATCTGATTCACATTAGCTTTCTGTAGACAG 480
138 ..... 138
481 TGGGTAGTCTCTGGGGGAGCTGAATTTGTAAGGTTACTCGGAAGG 530
138 ..... 138
531 CGAAGGGGGGGGCTGAGGGAAGTAGAGATTGTACACTCTCTGCTC 580
139 .....ValIleMetGlnAlaPheSerArgGluAlaLeuGln 150
581 CTGGGGGGTGTCAAGTGTATATGACAGGCTTCAGCCGGGAGCACTCA 630
150 nCysTrpValLeuValIleAlaGluGluValSerSerCysLeuGlnInt 167
|||||
631 GGTCTACGTGGCCCGTGCATCGCTGAGGAAGTCAAGATTCTCGGAGCA 680
167 rPLeuSerCysGlyGluArgGlyLeuLeuValTrpProGluValLysArg 183
|||||
681 GGCTTAAGTGGCGGCGAGCGCGCTCTCTGTCTACCCGAGGTGAAGCGC 730
184 leuMetPheArgIleAlaMetArgIleLeuLeuGlyCysGluProGlyPr 200
|||||
731 CTCATGTTCCGATCGCATCGCATCTCTGTGGCTGGAGACCGGCTCC 780
200 cAlaGlyGlyGlyGluAspGluGlnGlnLeuValGluAlaPheGluGln 217
|||||
781 AGCGGGCGGGGGGAGAGACAGACAGCTGTGAGAGCTTTTCGAGAGA 830
217 eThrArgAsnLeuPheSerLeuProIleAspValProPheSerGlyLeu 233
|||||
831 TGACCGGCAATCTTCTCTCTCTCCATTTGACGTGCGCTTACGCGGCTG 880
234 TyrArg..... 235
|||||

```

881 TACCGGTAAAGGCGGTTTCCGAGTCGGAGTAGGGGAACGCAAGCTCGG 930
 286 Gly ValLysAlaArg 240
 931 GCATCCGCTACCCGCCAGCCTCTCTCCGCGCTCAGAGGCGTGAAGGCGG 980
 241 AsnLeuIleHisAlaArgIleGluGluAsnIleArgAlaLysIleArgAr 257
 981 AACCTTATACAGCGCGCATGAGAGAAACATTCCGCCAAGATCCGCGG 1030
 257 GluGluAlaThrGluProAspGlyGlyCysLysAspAlaLeuGlnLeuL 274
 1031 GCTTCAGGCTACAGAGCCGAGTGGGGGTTTGCAGAGCGCGCTGCAGCTCC 1080
 274 euIleGluHisSerTrpGluArgGlyIleArgLeuAspMetGln..... 288
 1081 TGATTGACCACTCTGGGAGAGGGAGAGAGGTGATATGCAAGGTGAGA 1130
 288 288
 1131 AGCAATTCAAAAGGTGCCAAGGCGCGGAGTGCCTCTGACTTTCAGA 1180
 288 288
 1181 CACACTTCTGGGGTCTCCAAAGCCCTGTCAAGGCCCGCAGTACTTCCAA 1230
 288 288
 1231 GTGGCGCGCATGCTAGTCTAGAGCTTTTCAACTGTGGGTCTGACCC 1280
 288 288
 1281 CTTACGAGGACCAACCAACCCCTTTCAGAAAGGTGCGCTAAGAGCATCTGC 1330
 288 288
 1331 ATATCCGATATTACATCAGAAACATACAGTACGCAAAATTACCGTTAT 1380
 288 288
 1381 GAAGTAGCAACAAAGATATTATTATCGTTGGGGTCCACCACACACGAG 1430
 288 288
 1431 AACCGTATTAAAGGTGCGCATTTGTCTAGAGAGCTGTGAAAGGGGTGCG 1480
 288 288
 1481 TGAGCAATGGGGAAGATCCCAAAGTTCAAAGGCAAGGCTCATCTACAA 1530
 288 288
 1531 GGTAAAGCGGAAGAGCAGATTAAGGAGTTTCCGTTTTTGTGTGG 1580
 288 288
 1581 TCTTGACTTCTATGAAACAAACGATTTTACCCTGAAGCTTCCGTG 1630
 288 288
 1631 CAATATTCTCAGTCCAGTCTTTGTACAGTGCATTAACCTGCACACTAGA 1680
 289 AlaLeuLysGlnSerSerThr 295
 1681 TCTGATTAACCTTCGTTTTTATCCTTTAGGCACATAAACATTCGTCACA 1730
 296 GluLeuLeuPheGlyGlyHisGluThrThrAlaSerAlaAlaThrSerLe 312
 1731 GAGTCCCTCTTGGTGTATGAAACTACAGCCAGTCTGCGACGCTCACT 1780
 312 uIleThrTyrLeuGlyLeuTyrProHisValLeuGlnLysValaArgGluG 329
 1781 GATCACTTACCTAGGACTTACCCACATGATCTCCAGAAAGTTCCGAAAG 1830

329 IuIleLysSer..... 332
 1831 AGATTAAGAGCAAGTAGATGATATTCTAGAGTTCGCCATTGCCCTAGAA 1880
 332 332
 1881 CATTCTCTATTAAACACCACCACCACCACCCCACTGATATATAGTTTGTCTC 1930
 332 332
 1931 GATACACCCAGTACTATGACAGTGAAGATGTAGAGCTAGTGGGACTGT 1980
 332 332
 1981 GGGGAGAGACTCCACCTCTGTAATTTAAAAAGCAGTTGTTGACTGG 2030
 332 332
 2031 GCTCTCTCTTGGGCAGAATTGACCCCTCTCCTCTCTCTCTCTCTCTCT 2080
 332 332
 2081 CCTCTCTCTCTCTCCACACACACCACCACACCTTTTATAGAGCAAG 2130
 332 332
 2131 TTCTCTCTTCCCTGACCAAGAACATGATATGTGATTAGAGCCAAATGC 2180
 332 332
 2181 TGATCAGGTCGACAGTGTGTGAGGGCTCAGGGTATGACCTTTATATA 2230
 332 332
 2231 CCTGATAGCAACATTTCTGTGATATGGTTTAGGCTGAGGAAGTGG 2280
 332 332
 2281 AAAGAGGCCATCAGGCGCATCAGCTCTTCCCTTTATCTCTCCCATC 2330
 332 332
 2331 CAGAGCGCTTCAAGTTTAGTTAACAGTGAAGTCTGTGGCGTACTTTT 2380
 333 Lys. 333
 2381 TTTTGAAGTCCAGGATCCATCCATCACTCTTTTATCTGTGTTCCATA 2430
 334 GlyLeuLeuCysLysSerAsnGlnAspAsnLysLeuAspMetGluThrL 350
 2431 GGGCTTACTTTGCAGAGCAATCAAGACACAGTATAGACATGGAAGACTT 2480
 350 euGluGlnLeuLysTyrIleGlyCysValIleLysGluThrLeuArgLeu 366
 2481 TGGCACAGCTTAAATATACACTGGGTGTGATTAAGAGACCTCGGATTTG 2530
 367 AsnProProValProGlyGlyPheArgValAlaLeuLysThrPheGlu.. 382
 2531 AATCTCCGGTTCAGAGAGGTTTGGGTTGCTCTGAAGACTTTTGAGCT 2580
 382 382
 2581 GAATGTAGTGACCTCTCTGTCCCCACCCAGCCCTGTCCACAGTCCA 2630
 382 382
 2631 CTCTGCTATGTGTGAGCATCAGCTGCCAGAGCAGTGGCTCACTGCC 2680
 382 382
 2681 TTGACAGTGTCTGCTCTCTATGTACTGGGAAACCAATTGCTCTCTCT 2730

```

382 ..... 382
2731 CTTAATGCCATCATGCTAGTAATGACTTTTGTGTTCAGACTCAGGG 2780
383 .....Leuasn..GlyTyrG 387
2781 CCGGGATTGTCATTCATTAGGATTTTATTTTATTTTAAACAGGGATACC 2830
387 InIleProLysGlyTTrpAsnValIleTyrSerIleCysAspThrHisAsp 403
2831 AGATCCCCAAGGCTGGAATGTATTACACTATCTGACACCCACGAT 2880
404 ValAlaAspIlePheThrAsnLysGluGluPheAsnProAspArgPheI 420
2881 GTGGCAGATATCTTCACTAACACAGAGGAATTAATCCCGACCGCTTAT 2930
420 eValProHisProGluAspAlaSerArgPheSerPheIleProPheGlyG 437
2931 AGTCCCTCATCCAGAGATGCTTCCCGGTTCCAGCTTCATTCATTGGAG 2980
437 IyGlyLeuArgSerCysValGlyLysGluPheAlaLysIleLeuLeuLys 453
2981 GAGCCCTTCGAGAGCTGTGTAGCAAAAGATTGCAAAAATTCCTTTAG 3030
454 IlePheThrValGluLeuAlaArgHisCysAspTrpGluLeuAsnGlu 470
3031 ATATTTCAGTGGAGCTGGCTAGGCACTGATGAGCAGCTTCTAATAGG 3080
470 yProProThrMetLysThrSerProThrValTyrProValAspAsnLeuP 487
3081 ACCGCTCATCATGAAAGACACACCCCACTGTGACCTGTGGACATTCGC 3130
487 roAlaArgPheThrTyrPheGlnGlyAspIle 497
3131 CTGCAAGATTACCCACTTCACAGGAGATATC 3162
seq_name: /cgn2_6/prodata/2/lna/6a_COMB.seq:us-08-724-466B-3
seq_documentation_block:
; Sequence 3, Application US/08724466B
; Patent No. 6063606
; GENERAL INFORMATION:
; APPLICANT: Petkovich, P. Martin, White, Jay A.,
; TITLE OF INVENTION: Retinoid Metabolizing Protein
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Blake, Cassels & Graydon
; STREET: Box 25, Commerce Court West
; CITY: Toronto
; ZIP: M5L 1A9
; COUNTRY: Canada
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 Mb storage
; OPERATING SYSTEM: MS-DOS 5.1
; SOFTWARE: WORD PERFECT
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/724,466B
; FILING DATE: October 1, 1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/667,546
; FILING DATE: June 21, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Hunt, John C.
; REGISTRATION NUMBER: 36,424
; REFERENCE/DOCKET NUMBER: 50767/00004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 863-4344
; TELEFAX: (416) 863-2653
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1850 base pairs

```

```

; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-724-466B-3

alignment_scores:
    Quality: 1734.00      Length: 496
    Ratio: 4.070          Gaps: 4
    Percent Similarity: 85.887      Percent Identity: 67.742

alignment_block:
US-09-668-482-32 x US-08-724-466B-3 ..
Align seg 1/1 to: US-08-724-466B-3 from: 1 to: 1850

1 MetGlyLeuProAlaLeuAlaSerAlaLeuCysThrPheValLeuPr 17
   ||||| ..... |||||
88 ATGGGGCTGACACCCCTTATGGTCACTTCTCTGCACCATGCTGTACC 137
17 OleuLeuPheLeuAlaIleAlaLysLeuTrpAspLeuTyrCysValS 34
   ||||| ..... |||||
138 GCTTTACTCTTCTGCGCGGTGAAGTTGTGGAGATGTTAATGATCC 187
34 eSerArgAspArgSerCysAlaLeuProLeuProGlyThrMetGly 50
   ||| ..... |||
188 GACGAGTCATCCGAACCTCAGAAAGCTCTACCGCGCAGATGACATGGGC 237
51 PheProPhePheGlyGluThrLeuGlnMetValLeuGlnArgArgLysPh 67
   ||||| ..... |||||
238 TTGCGCTTCATTTGAGAAACGCTCCAGCTGATCTCCAGAGAAAGAGTT 287
67 eLeuGlnMetLysArgArgLysTyrGlyPheIleTyrLysThrHisLeuP 84
   ||||| ..... |||||
288 TCTGGCATGAAACGGCAGAAATACGGGTGCATCTTCAAGACGACCTCT 337
84 heGlyArgProThrValArgValMetGlyAlaAspAsnValArgArgL 100
   ||||| ..... |||||
338 TGGGGAACCCGACTGTACAGGTGATGGAGTGATGATATGAGGCAGATT 387
101 LeuLeuGlyGluHisArgLeuValSerValHisTrpProAlaSerValNr 117
   ||||| ..... |||||
388 CTGCTGGGGAACACAAAGCTGTGTCTGTTCAGTGGCCAGCACTGAG 437
117 gThrIleLeuGlyAlaGlyCysLeuSerAsnLeuHisAspSerHisAl 134
   ||||| ..... |||||
438 AACCATCTGGGCTGTGACACCTCTCCAAATGTCATGGAGTTCACACA 487
134 yGlnArgLysLysValIleMetGlnAlaPheSerArgLysAlaLeuGln 150
   ||||| ..... |||||
488 AAACAAAGAAAGGCCATTATGAGGGCGTTCTCGAGATGCTGGAG 537
151 CysTyrValLeuValIleAlaGluValSerSerCysLeuGlnInTr 167
   ||||| ..... |||||
538 CACTACATTCCTCCGATCCAGCAGGAGGTGAAGGCGCATACAGAAATG 587
167 pleu ..... SerCysGlyLysArgLysLeuValTyrProGlyLys 181
   ||||| ..... |||||
588 GCTGCAAAAAGACTCCTGC ..... GTGCTGTTTATCCAGAAA 625
181 allysArgLeuMetPheArgIleAlaMetArgIleLeuLeuGlyCysGlu 197
   ||||| ..... |||||
626 TGAAGAAACTCATGTTCCGATAGCTATGAGATCCTGCTGTTTGA 675
198 ProGlyProAlaGlyGlyGlyLysAspGluGlnGluValGlnAlaPh 214
   ||||| ..... |||||
676 CCAGACCAATA ..... AAGACGGACGACCAAGAACTGTGGAAGCTT 719
214 eGluGluMetThrArgAsnLeuPheSerLeuProIleAspValProPheS 231
   ||||| ..... |||||
720 TGAAGAAATGATCAAAACTGTTGCTTCCCTTGCACATGAGATTCCTTCA 769
231 eArgLeuTyrArgGlyValLysAlaArgAsnLeuIleHisAlaArgIle 247

```



```

117 gthrlleleuglyAlaGlyCysLeuSerAsnLeuHisAspSerSerHisL 134
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
438 AACCATCTGGGCTGTGACACCCCTTCATGTCATGTCATGTCATGTCATG 487
134 ysglnatgylslyValIleMetGlnIaPheSerArgIuaIaIeGln 150
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
488 AAACCAAGAAAAAGGCATTATGAGGGCTTCTCGAATGCTGAG 537
151 CystyrValleuValIleAlaGluIValSerSerCysLeuGlnItr 167
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
538 CACATACATCCCGTCATCCAGACAGAGGTGACAGCCCATACAGAGATG 587
167 pLeu.....SerCysGlyIuArgGlyLeuLeuValItyrProGluV 181
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
588 GCTGCAAAAAGACTCTGTC.....GTGCTGTTTATCTCCAGAAA 625
181 allysArgLeuMetPheArgIleAlaMetArgIleLeuLeuGlyCysGlu 197
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
626 TGAAGAAACTATGTTTGGATAGCTATAGAAATCTGCTTGGTTTGA 675
198 ProGlyProAlaGlyGlyGlyGlyIuAspGluGlnIleuValGluAlaPh 214
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
676 CCAGAGCAAAATA.....AAGACGAGACAGACAGACACTGTCGAGGCTT 719
214 eglIuMetThrArgAsnLeuPheSerLeuProIleAspValProPheS 231
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
720 TGAGGAATGATCAAAAACCTGTTCTCTCTCCCAATCGACCTCTCTTCA 769
231 eglIuLeuTyrArgGlyValIlysAlaArgAsnLeuIleHisAlaArgIle 247
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
770 GGGCTGTGACGGGGTTTGAGGGACCAATTCATTCATCTCCAAAT 819
248 GluIuAsnIleArgAlaIlysIleArgArgLeuGlnAlaItrGluProAs 264
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
820 GAGGAACATCAGAGAGAAATCAACAT...GACGACATGAAAACGA 866
264 pGlyGlyCysIysAspAlaLeuGlnLeuLeuIleGluHisSerTrpGluA 281
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
867 ACAGAAATACAAAGACGCCCTTACCTGTGTCGAGAACGACAGAA 916
281 rglGlyIuArgLeuAspMetGlnAlaLeuIysGlnSerSerThrGluLeu 297
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
917 GTGACGAACTTTAGTTTGCAGGGCATGAAAGAACGACATACAGCTT 966
298 LeuPheGlyGlyHisGluThrThrAlaSerAlaIaItrSerLeuIleth 314
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
967 CATATTGGAGTCATGAAACCCAGCCAGACACTGCACCTCATTGTCAT 1016
314 rTyIleuGlyLeuTyrProHisValLeuGlnIuValArgGluGluIleL 331
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1017 GTTTCGGGCTGGAACACAGAGGTGCGAAGGTGCAAGAGAGGCTTC 1066
331 ySerIysGlyLeuLeuCysIysSerAsnGlnAspAsnIysLeuAspMet 347
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1067 AGCAGAGAGTTGAATGGCATGTATACACCGCGAAAGGCTTGGATAG 1116
348 GluThrLeuGluGlnLeuIysTyrIleGlyCysValIleIysGluThrIe 364
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1117 GAGCTGTGTGACAGCTGAGTACACTGATGTGTGATTAAGAGACTCT 1166
364 uArgLeuAsnProProValProGlyIuPheArgValAlaIaIeIysThrP 381
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1167 TAGAATCAACCTCTCTGTCCTCCGAGAGATTCAGAGTCGACACCAAACT 1216
381 heglIuLeuAsnGlyTyrGlnIleProIysGlyTyrAsnValIleItyrSer 397
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1217 TTGAATTGAATGTTTACCAAAATTCCTAAAGATGAGAACTCATTTACAGC 1266
398 IleCysAspThrHisAspValAlaAspIlePheThrAsnIysGlnIuPh 414
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1267 ATCTGTGACACGACGATGTGGCGAGCTTTCACAAACAGAGAGATT 1316

```

```

414 eAsnProAspArgPheIleValProHisProGluAspAlaSerArgPheS 431
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1317 CCAGCCGGAGATTCATGACAAAGTGTGAGAGACGGGTCCAGGTTTA 1366
431 erPheIleProPheGlyGlyGlyLeuArgSerCysValGlyIysGluPhe 447
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1367 ACTACATCCCTTCGAGAGAGATCCAGATGTGTGTGGCAAGAGCTTC 1416
448 AlaIysIleLeuLeuIysIlePheThrValGluLeuAlaArgHisCysAs 464
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1417 GCCAAGGTACTCAAGATCTTTAGTTGAGTTACCGACGATTCGCA 1466
464 pTrpGlnLeuLeuAsnGlyProProThrMetIysThrSerProThrValT 481
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1467 TTGGATTTCTCAACGAGACCCCGACAAATGAAACAGGCCGACTATT 1516
1517 ACCCAGTGACAATCTCCACCAAAATTCATGATGAT 1554
seq_name: /cgn2_6/plodata/2/lna/6B_COMB.seq:US-08-882-164D-36

```

```
seq_documentation_block:
; Sequence 36, Application US/08882164D
; Patent No. 6306624

```

```
GENERAL INFORMATION:

```

```

APPLICANT: Petkovich, P. Martin, White, Jay A.
APPLICANT: Beckett, Barbara R., Jones, Glenville
TITLE OF INVENTION: Retinoid Metabolizing Protein
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESS: Blake, Cassels & Graydon
STREET: Box 25, Commerce Court West
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5L 1A9

```

```
COMPUTER READABLE FORM:

```

```

MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 MB storage
COMPUTER: COMPAQ, IBM PC compatible
OPERATING SYSTEM: MS-DOS 5.1
SOFTWARE: WORD PERFECT

```

```
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/882,164D
FILING DATE: June 25, 1997

```

```
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/667,546
FILING DATE: June 21, 1996
APPLICATION NUMBER: 08/724,466
FILING DATE: October 1, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Hunt, John C.
REGISTRATION NUMBER: 36,424
REFERENCE/DOCKET NUMBER: 50767/00010
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 863-4344
TELEFAX: (416) 863-2653
INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 2677 base pairs
type: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-882-164D-36

```

```
US-08-882-164D-36

```

```

alignment_scores:
Quality: 1251.00      Length: 669
Ratio: 4.075          Gaps: 7
Percent Similarity: 45.889      Percent Identity: 43.797

```

```

alignment_block:
US-09-668-482-32 x US-08-882-164D-36

```

Align seg 1/1 to: US-08-882-164D-36 from: 1 to: 2677

```

1 MetGlyLeuProAlaLeuLeuAlaSerAlaLeuCysThrPheValLeuPr 17
676 ATGGGGCTCCCGCGCTGCTGGCCAGTGGCGCTCTCCACCTTCGCTGCG 725
17 OleuLeuPheLeuAlaAlaLeuLysLeuTrpAspLeuTyrCysValS 34
726 GCTGCTGCTCTCCGCGCTGGCATCAGCTCTGGAGCCGTACTGCTGCA 775
34 erSerArgAspArgSerCysAlaLeuProLeuProGlyThrMetGly 50
776 GCGGCGCGACCGCGAGTTGTGCTCCCTCCATGCCCCCGGAGCTATSGS 825
51 PheProPheGlyLysThrLeuGlnMetValLeu Gln..... 63
826 TTCCTCTCTTTGGGGAAACCTTGCAGATGNTACTMCAGGTAAAGGAGG 875
63 ..... 63
876 TGGGGCGGAGACAGCTGCTTCCCGGAGCCCGGCGGCTCTGGGCTTCT 925
63 ..... 63
926 GCTGAGAGTGGGGTAGGCGCCCGGGAGGAGCATGCTATTGCGCTAGAG 975
63 ..... 63
976 CAGGGCTGGGGGAGCGGGCGCTCCCGGKMYKSCICAMGCSRCKMWK 1025
64 .....ArgArgLysPheLeuGlnMetLysArg 72
1026 TMMCTCCGCTTCCTCCCAAGGCGGARSARMKCYKGRGATGAAGCGC 1075
73 ArgLysTyrGlyPheIleTyrLysThrHisLeuPheGlyArgProThy 89
1076 AGGAAATACGGCTTCATCTACAAAGCAGCATGTGTGGGCGGCCACCGT 1125
89 ArgValMetGlyAlaAspAsnValArgArgLysLeuLeuGlyLysHis 106
1126 ACGGCTGATGGCGCGGCAATGTGCGCGCATCTTGCTCGAGAGCACCC 1175
106 ArgLeuValSerValHisTrpProAlaSerValArgThrIleLeuGlyAla 122
1176 GCGTGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1225
123 GlyCysLeuSerAsnLeuHisAspSerSerHisLysGlnArgLysLys 138
1226 GGCTGGCTCTTAACCTGCAAGACTCTCTCCACAGCAGCGCAAGAAAGT 1275
138 ..... 138
1276 GGGGGAGAGGCGGAGCGCTGACAGGAGGAGGAGGCCATTATAGAGC 1325
138 ..... 138
1326 GAATTCGGCTGATGATGCTAGGCGGCGCTAGCAGCTTAGAGTGGCT 1375
138 ..... 138
1376 AGGACCTCTGCGCAGCTCCAGTTAGCTTCCAGCTCGAGAGAGTGCAT 1425
138 ..... 138
1426 GTGCTGCGAGGAGCTGGGGTGTGTGGAAGGAGGAGCGGCTAGACGAG 1475
139 .....ValIle 140
1476 GGGCGGATGAGGCTTTAAAGCTGTCCCTCTCGGAGCTCAGGATGAT 1525
141 MetGlnAlaPheSerArgGlnAlaLeuGlnCysTyrValLeuValIleAl 157
||||:||||:||||:||||:||||:||||:||||:||||:||||:

```

```

1526 ATGGGGCTTCAGCCGCGAGCACTCGAATGCTACGTCCGCGTATCAC 1575
157 agLgluValSerSerCysLysGlnIleTrpLeuSerCysGlyLysArg 174
||||:||||:||||:||||:||||:||||:||||:||||:||||:
1576 CGAGGAATGGGCGAGCAGCTCGAGACAGTGGTGAAGCTCGGCGAGCGG 1625
174 LysLeuValTyrProGlnValLysArgLeuMetPheArgIleAlaMet 190
1626 GCCCTCTGATACCCCGAGGTGAAGCCGCTCATGTTCGAATCGCATG 1675
191 ArgIleLeuLeuGlyCysGlnProGlyProAlaGlyLysGlyLysAsp 207
1676 CGCATCTACTGGCTGGCAACCCCACTGGCGCGAGCGGAGCTCCGA 1725
207 GlnGlnLeuValGlnAlaPheGlnGlnMetThrArgAsnLeuPheSer 224
1726 GCAGCAGCTTGTGGAGGCTTCGAGGAAATGACCGCAATCTCTCGC 1775
224 euProIleAspValProPheSerGlyLeuTyrArg..... 235
1776 TGCCATGACGTGCCCTTCAGCGGGGTGTACGGGTAAAGGCGGCAAC 1825
235 ..... 235
1826 GGGCTGCGGACTAGGGGCGCGGAGCTGGGCGTCTGCTACCGCGCGCG 1875
236 .....Gly.ValLysAlaArgAsnLeuIleHisAlaArgIle 247
1876 CTCTCTGGCTCAGGGGATGAAGGCGCGAACCCTCATTCACGCGCGCATC 1925
248 GlnGlnAsnIleArgAlaLysIleArgArgLeuGlnAlaTrpProAs 264
1926 GAGCAGAACATTCGGCGCAAGATCGCGGGCTGGCGGAGATCCAGGCGGG 1975
264 pGlyGlyCysLysAspAlaLeuGlnLeuLeuIleGlnHisSerTrpGln 281
1976 CCAAGGCTGCAGAAAGCGGCTGCAGCTGTGATGAGCAGCATCTGGGAGA 2025
281 ArgLysLysArgLeuAspMetGln..... 288
2026 GGGGAGAGCGGCTGGACATGCAAGGTGAGTACGACTTCAGACAGGCACT 2075
288 ..... 288
2076 GCGGAGTTTGTCCTCCGCTGGCTTCCAAAGCGCTGTCTCTGGGGCCCCAA 2125
288 ..... 288
2126 AGCGCGGCTGGGGCCAGCTTCTGAGTGGGCGCGGCTCAGACTA 2175
288 ..... 288
2176 CAGCTATGGAATCCGGAAGAGCTGAGACACCGGCTCAGAGAGCTGC 2225
288 ..... 288
2226 GGAAGGGGCTGCGGAGAACTGGAGACATCCCTACCTTTAMCAGGTT 2275
288 ..... 288
2276 TCAAGGGAAGTTGGAATTTGCAAAAATGTTAATAAAGACTTGGAT 2325
288 ..... 288
2326 TTTAATAAAGTAAGACTTTAAGTCAAGAGATTTCGGTAGCGGGGTCG 2375
288 ..... 288
2376 TACTCGCTTACTGTCACAGCTGAAGGAGGAGCTTGCATTTGTTTA 2425
288 ..... 288
2426 AAGATATGCTTCTGACTTCTGTGACGAAACATTTAGCCCTTCTA 2475

```



```
288 ..... 288
2476 GTCTCCCTCCAGAACTCTCAGTTCTGATTCTTCTGTCTCAAA 2525
289 .....Ala..Leu 290
2526 CCGCAGCAGCACTGTGTGAATGTGGCTCTCAGCTTATCTTAGCCACTA 2575
291 LysGlnSerSerThrGluLeuLeuPheGlyGlyHisGluThrAlaSer 307
2576 AAGCAATTTTCAACGAACTCTCTTTGGAGACACGAAACACAGCCGACG 2625
307 TALAAlaThrSerLeuIleuThrTyrLeuGlyLeuTyrProHisValLeuG 324
2626 TGCAGCCACATCTCTGATCACTTACCTGGGGCTCTACCCACATGTTCTCC 2675
324 In 324
2676 AG 2677

seq_name: /cgn2_6/prodata/2/1na/6A_COMB.seq:US-08-724-466B-11

seq_documentation_block:
; Sequence 11, Application US/08724466B
; Patent No. 6063606
; GENERAL INFORMATION:
; APPLICANT: Petkovich, P. Martin, White, Jay A.,
; APPLICANT: Beckett, Barbara R., Jones, Glenville
; TITLE OF INVENTION: Retinoid Metabolizing Protein
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Blake, Cassels & Graydon
; STREET: Box 25, Commerce Court West
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5L 1A9
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 Mb storage
; OPERATING SYSTEM: MS-DOS 5.1
; SOFTWARE: WORD PERFECT
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/724,466B
; FILING DATE: October 1, 1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/667,546
; FILING DATE: June 21, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Hunt, John C.
; REGISTRATION NUMBER: 36,424
; REFERENCE/DOCKET NUMBER: 50767/00004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 863-4344
; TELEFAX: (416) 863-2653
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 351 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-724-466B-11

alignment_scores:
Quality: 604.00 Length: 117
Ratio: 5.207 Gaps: 0
Percent Similarity: 99.145 Percent Identity: 97.436

alignment_block:
US-09-668-482-32 x US-08-724-466B-11 ..
Align seg 1/1 to: US-08-724-466B-11 from: 1 to: 351
```

```
296 GluLeuLeuPheGlyGlyHisGluThrAlaSerAlaAlaThrSerLe 312
1 GAACTCTCTTTGGAGACACGAAACACAGCCAGTGCACCATCTCT 50
312 uIleThrTyrLeuGlyLeuTyrProHisValLeuGlnHisValArgGluG 329
51 CATCACTTACCTGGGGCTCTTACCACATGTTCTCCAGAAAGTGGCAGAAG 100
329 LuIleLysSerLysGlyLeuLeuCysLysSerAsnGlnAspAsnLysLeu 345
101 AGCTGAAGAGTGAAGGGTTTACTTTCACAGCAATCAACAAACAAAGTTG 150
346 AspMetGluThrLeuGlnGlnLeuLysTyrIleGlyCysValIleLysG 362
151 GACATGGAAATTTTGGAACACTTAATATCATCGGGTGTGATTATTAAGA 200
362 uThrLeuArgLeuAsnProProValProGlyGlyPheArgValAlaLeuL 379
201 GACCCCTTGACATGAATCCCCCAGTTCCAGAGGGTTCCGGGTTGCTCTGA 250
379 ySThrPheGluLeuAsnGlyTyrGlnIleProLysGlyTyrPAsnValIle 395
251 AGACTTTTGAATTAATGATACACAGATTCCAAAGGGCTGAATGTATC 300
396 TyrSerIleCysAspThrHisAspValAlaAspIlePheThrAsnLysG 412
301 TACAGTATCTGTGATCATGATGATGTGGCAGAGATCTTCACCAACAAGA 350
412 u 412
351 A 351
```

```
seq_name: /cgn2_6/prodata/2/1na/6B_COMB.seq:US-08-882-164D-11

seq_documentation_block:
; Sequence 11, Application US/08882164D
; Patent No. 6306624
; GENERAL INFORMATION:
; APPLICANT: Petkovich, P. Martin, White, Jay A.,
; APPLICANT: Beckett, Barbara R., Jones, Glenville
; TITLE OF INVENTION: Retinoid Metabolizing Protein
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Blake, Cassels & Graydon
; STREET: Box 25, Commerce Court West
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5L 1A9
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 Mb storage
; OPERATING SYSTEM: MS-DOS 5.1
; SOFTWARE: WORD PERFECT
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/882,164D
; FILING DATE: June 25, 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/667,546
; FILING DATE: June 21, 1996
; APPLICATION NUMBER: 08/724,466
; FILING DATE: October 1, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Hunt, John C.
; REGISTRATION NUMBER: 36,424
; REFERENCE/DOCKET NUMBER: 50767/00010
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 863-4344
; TELEFAX: (416) 863-2653
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 351 base pairs
```

TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-882-164D-11

alignment_scores:
Quality: 604.00 Length: 117
Ratio: 5.207 Gaps: 0
Percent Similarity: 99.145 Percent Identity: 97.436

alignment_block:
US-09-668-482-32 x US-08-882-164D-11 ..

Align seg 1/1 to: US-08-882-164D-11 from: 1 to: 351

```
296 GluLeuLeuPheGlyGlyHisGluThrHisAlaSerAlaIleThSerle 312
|||||
1 GAACCTCTCTTTGGAGACACGAAACACGGCCAGTGCACACATCTCT 50
312 uLeuThrLeuGlyLeuTyProHisValLeuGlnLysValArgGlu 329
|||||
51 GATCCTTACCTGGGCTCTACCCACATGTTCTCCAGAAAGTGCAGAG 100
|||||
329 LuLeuLysSerLysGlyLeuLeuCysLysSerAsnGlnAspAsnLysLeu 345
|||||
101 AGCTGAGAGAGTAAGGTTACTTTCAGAGACATCAAGACAAAGTTG 150
|||||
346 AspMetGluThrLeuGluGlnLeuLysTyrlIleGlyCysValIleLysG 362
|||||
151 GACATGGAAATTTGGAAACACTTAATACATCGGGTGTATTATTAAGGA 200
|||||
362 uThrLeuArgLeuAsnProProValProGlyGlyPheArgValAlaLeuL 379
|||||
201 GACCTTGACTGATGATCCCGAGTCCAGAGAGGTTGGGTTGCTCTGA 250
|||||
379 yStrPheGluLeuAsnGlyTyrlGlnIleProLysGlyTyrPAsnValIle 395
|||||
251 AGACTTTTGAATTAATGATACCGATTCACAAAGGCTGGAATGTATTC 300
|||||
396 TyrSerIleCysAspThrHisAspValAlaAspIlePheThrAsnLysG 412
|||||
301 TACAGTATCTGTGATACATCATGTGTGCGACAGATCTTCCACCAACAAGA 350
|||||
412 u 412
351 A 351
```

seq_name: /cgn2_6/prodata/2/lna/5B_COMB.seq:US-08-622-166A-1

seq_documentation_block:

Sequence 1, Application US/08622166A
Patent No. 5952545
GENERAL INFORMATION:
APPLICANT: KONCZ, CSABA
APPLICANT: MATTHUR, JAIDEEP
APPLICANT: SEKERES, MIKLOS
APPLICANT: ALTMANN, THOMAS
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING CYTOCHROME
NUMBER OF INVENTION: P450-TYPE PROTEINS INVOLVED IN THE BRASSINOSTEROID SYNTHESIS
CORRESPONDENCE ADDRESS:
ADDRESSEE: Birch, Stewart, Kolasch & Birch, LLP
STREET: P.O. Box 747
CITY: Falls Church
STATE: Virginia
COUNTRY: USA
ZIP: 22040-0747
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/622,166A
FILING DATE: 27-MAR-1996
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: SVENSSON, LEONARD R.
REGISTRATION NUMBER: 30,330
REFERENCE/DOCKET NUMBER: 0147-0153P
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 205-8000
TELEFAX: (703) 205-8050
TELEX: 248345
INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:
LENGTH: 1608 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO

ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: A. thaliana
IMMEDIATE SOURCE:
LIBRARY: lambda gt10
CLONE: C204

FEATURE:
NAME/KEY: CDS
LOCATION: 48..1466
US-08-622-166A-1

alignment_scores:
Quality: 450.00 Length: 503
Ratio: 1.562 Gaps: 15
Percent Similarity: 57.256 Percent Identity: 26.640

alignment_block:
US-09-668-482-32 x US-08-622-166A-1 ..

Align seg 1/1 to: US-08-622-166A-1 from: 1 to: 1608

```
6 LeuLeuAlaSerAlaLeuCysThrPheValLeuProLeuLeuPheLe 22
|||||
72 CTCCTCTCTCCATCGCCGCGCTTCTCCTCTACTC..... 110
|||||
22 uAlaAlaLeuLysLeuTrpAspLeuTyrCysValSerSerArgAspArg 39
|||||
111 .....CGCGGTACACGTT 123
|||||
39 erCysAlaLeuProLeuProGlyThrMetGlyPheProPhePheGly 55
|||||
124 ACCGTCGATGGGTCTGCTCCGGAAGCTTGGTCTCTCTGATAGGA 173
|||||
56 GluThrLeuGlnMetVal.....LeuGlnArgArgLysPhe 67
|||||
174 GAGACTTTTCAGTCGATCGAGCTTACAAACAGAGAACCTGAGCCTT 223
|||||
67 eLeuGlnMetLysArgArgLysTyrlGlyPheIleTyrlLysThrHisLeu 84
|||||
224 CATGACGACAGAGAGTACGCCGCTGAGTCTGCTTTCATGACGAGATCTT 273
|||||
84 heGlyArgProThrValArgValMetGlyAlaAspAsnValArgArgIle 100
|||||
274 TTGCTGAACCGACATTTTTCACACTGACCCGGAACGACGCTTTGTT 323
|||||
101 LeuLeuGlyGlnHisArgLeuValSerValHisTrpProAlaSerValAr 117
|||||
324 CTCGAGAACGAGAGGAGCTTTTTCAGTGTCTTATCTGCTCTCAATTTG 373
|||||
117 gThrIleLeuGlyAlaGlyCysLeuSerAsnLeuHisAspSerGlnSL 134
|||||
374 TACCTTTTGGGAAACACTCTCTTCTTATGAAGGTTCTTTGGATA 423
```



```
80 GTGTCCTCTTTTCATCTTCATCAGC..... 106
31 rcysvalserSerArgspArSerSerCysAlaLeuProLeuProPro147 48
      ::::: ||| ||| ||| ||| ||| |||
107 .TTCATCAGACGGGGGAGGCGCT.....CCATATCTCTCCGCGTC 146
48 hrMetGlyPheProPhePheGlyLThrLeuGlnMetValLeuGlnArg 64
      |||:::||||:::||||:::||||:::
147 CACGAGTTGGCCATCATAGGCAACATGTTAATGATGACCAACTCACC 196
65 ArgLysPheLeuGlnMetLysArgArgLysTyrGlyPheLeuTyrLys 81
      ::::: ||| ::::: ||| ||| ||| |||
197 CACCGTGGTTAGCCCAATTAGCTAAAGATGCGGAGTTG.....TG 240
81 rHisLeu...PheGlyArgProThrValArgValMetGlyAlaAspAsn 97
      ||| ||| ||| ||| ::::: ||| ::::: |||
241 CCATCTCCGCGCATGGGATTCCTCCATATGTACGCTGCTCATACCCAGG 290
97 aArgArgLLeuLeuGlyLThrLysArgLeuValSerValHisTrpPro 113
      || ||| ::::: ||| ::::: |||
291 TGGCTCGACAGTCCCTTCATAGTCCAGACAGCGCTTCTCGACCGGCT 340
114 AlaSerValArgThrLLeuGlyAlaGlyCysLeuSerAsnLeu...H1 129
      |||:::||||:::||||:::||||:::
341 GCACACTATA.....GCTATTAAGCTATCTGACTTA 369
129 sasPserSer.....HisLysG 135
      ::::: ||| ::::: |||
370 CGACCGACGCGACATGGCTTCGCTCCACTACGAGCCGTTTGGAGACAGA 419
135 InArgLysLysValLLeuMetGlnAlaPheSerArgGluAlaLeuGln 151
      ::::: ||| ::::: ||| ||| |||
420 TGAGAAAGTGTGTGTATGAGGTGTTAGCCGTTAAAGAGCTGACTCA 469
152 TyrValLLeuVal.....LLeuLagLugLValSerSe 162
      ::::: ||| ||| ||| ||| ||| |||
470 TGGGCTTACGTTCCGTGATGAGAGTGCAGAAATGTCGCGTCTTCTTG 519
162 rcysLeuGlnGlnTrpLeuSerCysGlyLThrArgLysLeuValTyrP 179
      ::::: ||| ::::: ||| ||| |||
520 TACGTTGGTAGCCTATTAACGTGGGGAGACA.....ATTWTG 560
179 rcGluValLysArgLeuMetPheArgLLeuMetArgLLeuLeuGly 195
      ::::: ||| ::::: ||| ||| |||
561 CACTGACCGCGCAACATTAACCGGCAAGCGTTGG.....TCACCC 604
196 CysGluProGlyProAlaGlyLysGluAspGlnGlnLeuValG1 212
      ||| ||| ||| ::::: |||
605 TCCGAGACAGGA.....CAAGACGATTCATAG 633
212 uAlaPheGlnGlnMetThrArg.....AsnLeuPheSerLeuProIleA 227
      ::::: ||| ::::: ||| ||| |||
634 AATCTTACAGAGTCTCTTAAGCTTTTGGACCTTCAACGTAAGCGGAT 683
227 spValProPheSerGlyLeuTyr.....ArgGly 236
      ::::: ||| ||| ||| ||| |||
684 TCATACCATAATTTCGGGTGATCGATCCGCAAGGATAAACAAGCGGCTC 733
237 ValLysAlaArgAsnLeuLLeuLLeuLLeuLLeuLLeuLLeuLLeu 251
      ||| ||| ||| ||| ::::: ||| ||| |||
734 GTAAAGCCCGTAATGATCTAGACGATTTATGACGATATATCATGATGA 783
252 .ArgAlaLysLLeuArgLeuGlnAlaThrGluProAspGlyGly... 266
      ::::: ||| ::::: ||| ||| ||| |||
784 ACATATGAAGAGAGAGATCAAAACGCTGTGATGGGAGATGTG 833
267 .....CysLysAspAlaLeuGlnLeuLeuLeuLeuLeuLeuLeu 278
      ||| ||| ||| ||| ||| ||| |||
834 TCGATACCGGATATGTTGATGATCTTCTTGTCTTTTACAGTGAAGGCC 883
279 TrpGluArgGlyLysGluArgLeuAspMetGln..... 288
      ::::: ||| ||| ||| ||| ||| |||
884 AATTTAGTCAGTGAGACGCGGATCTCAAAATTCATCAAACTTACCGG 933
```

```
289 ...AlaLeuLysGlnSerSerThrGlnLeuLeuPheGlyLysGlnT 304
      ::::: ||| ||| ||| ||| ||| |||
934 TGACAAATATCAAGCAATCATATGACGCTTATGTTGGAGACAGGAAA 983
304 hrThrAlaSerAlaAlaThrSerLeuLLeuThrTyrLeuGlyLeuTyrPro 320
      |||:::|||||||:::|||||||
984 CGGTACCGTCCGGGATAGAGTGGCGCTTAACGAGATTATTAAGGAGCC 1033
321 HisValLeuGlnLysValArgGluGluLe...LysSerLysGlyLeuLe 336
      ::::: ||| ::::: ||| ||| ||| |||
1034 GAGAGCTAAAGACGGGTCCAAACAGAACTCCGCAAGTCTTGACTT... 1081
336 ucLysLysSerAsnGlnAspAsnLysLeuAspMetGluThrLeuGln 353
      |||:::|||||||:::|||||||
1082 .....GACGAGAGGTGAAGATCCGACATCGAGAAAT 1115
353 eutLysTyrLLeuGlyCysValLLeuGlyLThrLeuArgLeuAsnPro 369
      || |||:::|||||||:::|||||||
1116 TGACTTATCTCAATGACACCTCAAGAAACCTTAAGATGACACCCAGG 1165
370 ValProGlyLysPheArgValAlaLeuLysThrPheGluLeuAsnLys 386
      ::::: ||| ::::: ||| ||| ||| |||
1166 ATCCCTCTCTCTCCACGAACCGCGGAGACACTAGTATCGACGCTT 1215
386 rGlnLLeuProLysGlyTyrAsnValLLeuTyrSerLLeuCysAspThr 403
      ::::: ||| ||| ||| ||| ||| |||
1216 CTTCATTCGCCAAGATCTGCTGATGATCAACGCGTTGCCATAGAGC 1265
403 spValAlaAspLLeuPheThrAsnLysGlnLThrPheAsnProAspArg 419
      ::::: ||| ||| ||| ||| ||| |||
1266 CGGACCAACCTCTTGACATGACCGGACAGCTTATACCATGAGGTTT 1315
420 LLeuValProHisProGluAsp.....AlaSerArgPheSerLLeu 434
      ::::: ||| ||| ||| ||| ||| |||
1316 TTGGAAACCGGCGCTACCGGATTCMAAGGAGCAATTCAGATTATAC 1365
434 oPheGlyLysGlyLeuArgSerCysValGlyLysGluPheAlaLysL 451
      ||| ||| ||| ||| ||| ||| |||
1366 GTTGGGGCGGGTGTGATGATGTCGCCGGGATGCA..... 1402
451 euleuLysLLeuPheThrValGluLeuAla.....ArgHisCys 463
      ||| ::::: ||| ||| ||| ||| |||
1403 ..CTAGGTTATACCGCTGACTTACCGCGGTCTCATATATTCATTGC 1450
464 ...AspTrpGlnLeuLeuAsnGlyProProThrMetLysThrSer 477
      |||:::|||||||:::|||||||
1451 TTCACGTGGAATTA.....CTGATGGGATGAACCAAGT 1486
seq_name: /cgn2_6/prodata/2/lna/6B_COMB.seq:us-09-103-840A-2
seq_documentation_block:
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: PRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
```

OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

alignment_scores:

Quality:	329.00	Length:	456
Ratio:	1.348	Gaps:	12
Percent Similarity:	53.509	Percent Identity:	24.123

```
alignment_block: 72 00 100 0400 0 64
```

Align seg 1/1 to reverse of: US-09-103-840A-2 from: 1 to: 4403765

32 CysValSerSerArgAspArgSerCysAlaLeuProLeuProProGlyH 48
||| ||| ||| : : : : : ||| |||
4120837TGCAGCTCCCGGACGACGAGCGCGGATCTTTGGCGGCCACC GGCT 4120788

48 rMetGlyPheProPhePheGlyGluThrLeuGlnMetValLeuGln... 63
||| ||| : : : : : ||| : : : : :
4120787TGCAGCGCGCCCGTGATTGTGCGCGCGCTTCGCGCGCTACGCG GTT 4120738

64 Arg.ArgLysPheLeuGlnMetLysArgGly 74
||| ||| : : : : : ||| : : : : :
4120737GCCTACCTGCGCGGTGGCGAGCGCCGATGCTGCAGACTGATCGT CG.. 4120690

74 sTyrgLyPheIle..... 78
||||| |||
4120689..TACGAGCATGATCAAGACGATCCCATTCTCACGCTGGTAGCTTCGA 4120641

79 TyrLysThrHisLeuPheGlyArgProThrVal 89
: : : : : ||| |||
4120640TGCATTCCGACACGCGCGCTGGCCAAAGAGGTTTTACACCGCTCCACCGCAC 4120591

90 ArgValMetGlyAlaAspAsnValArgArgIleLeuLeuGlyLuhisAr 106
: : : : : |||
4120590GCTCCCTCTCGGGGAGAGGGGGT..... 4120567

106 gLeuValSerValHisTrpProAlaSerValArgThrIleLeuGlyAlaG 123
||||| : : : : : ||| : : : : :
4120566.....GGGCGGGACGCG....GGCATCAATAGCGTCGG 4120539

123 LcYsLeuSerAsnLeuHisAspSerSerHisLysGlnArgLysLysVal 139
|| : : : : : : : : : : : ||| : : : : : ||| : : : : :
4120538GGTCGATGTTTCTCCAGAGAGAGCGCGGACATCTTCACGCGCAAATC 4120489

140 IleMetGlnAlaPheSerArgGlnAlaLeuGlnCysTryValLeuValII 156
: : : : : ||| : : : : : ||| : : : : : ||| : : : : :
4120488TTGACGCGCCCATATCAATGAGCGGCTCGCTGATGATTACGTTCCGAT TAT 4120439

156 eAlaGluGluValSerSerCysLeuGluGlnTrpLeuSerCysLysLya 173
: : : : : ||| : : : : : ||| : : : : : ||| : : : : :
4120438CGAAGATCTCACTCCGCGGCGGATGCAATCAATCAACCGCGTT....GACC 4120395

173 rgGlyLeuLeuValIlyrProGluValIlysArgLeuMetPheArgIleAla 189
|| : : : : : : : : : : : ||| : : : : : : : : : : :
4120394AGCTTTGCCATGTTCAGCGGCGCGGTGCGATGCTGAGATGATATC 4120345

190 MetArgIleLeuGlyCysGluProGlyProIleArgIlyGlyLys 206
: : : : : ||| : : : : : ||| : : : : : ||| : : : : :
4120344GTCAAGGTCATATTGGGGTGCAC.....GACCCGAGGA 4120310

206 pGluGlnGlnLeuValGIuaIAPheGluGluMetThrArgAsnLeuPheS 223
: : : : : ||| : : : : : ||| : : : : : : : : : : :
4120309GGTCGCGGCTTGGCGCGCGCTTCGACGCGCTGTAAATCTGGTGT TT 4120260

223 erLeuProIleAspAlaProPheSer.....GlyLeuTyrArg 235
|| : : : : : : : : : : : ||| : : : : : ||| : : : : :
4120259CAGAGCATTAACGCTGCGTACGACGACATCGCTGGCGGACACTACA GA 4120210

236 GlyValValArgAsnLeuIleHisAlaArgIleGluGluAsnIleAr 252
||||| : : : : : ||| : : : : : ||| : : : : :
4120209GTCCTGCCCGACGCGCCCGCCCAACGAGATGACGACGTCGTCAT 4120160

```

252   GAlAlaYsIlLeArGArGleuGlnAlaThcIuPrOaSpRlYglYcYsYsA 269
      ||| |||   :   :   :   :   :
4120159GCGCTCAATCCGAGCGGGCGCGCATCCACCGCTGGGAGCAGACG 4120110

269   sPaLaLeuGlnLeuLeuIlleGlnHSerTrrPgluArGlygluArGleu 285
      :   :   :   :   :   :   :   :   :   :   :   :   :
4120109ACGTCGTGTGGTTCGTGAGCGCGCGGAGGACATCCGGGAAACATGG 4120060

286   AsPmeGlnAlaLeuYsGlnSerTrrHgluLeuLeuPheGlygluH 302
      :   :   :   :   :   :   :   :   :   :   :   :   :
4120059TCAGACGCGAGATCCGCGCGCATCTGATCACCGCTGCGTGGGGGAG 4120010

302   sGluTrrHrAlaSerAlaAlaHrSerLeuHrYrLeuHrYrLeuT 319
      |||||TrrHr|||||:   :   :   :   :   :   :
4120009GTAAACACGCGCGCACCGTGGCGTGGCGCTTACACTTGGCTGCAC 4119960

319   YrrPrOHsValLeuGlnLysValArGluGlnIlleYsSerTrrGlyleu 335
      :   :   :   :   :   :   :   :   :   :   :   :   :
4119959ATCGGAGCGCTTGGCGCGCTGCGAGCGAGCGGCTGGCGTGGG 4119913

336   LeuYsYsSerAsnGlnAsPaSnLysLeuAsPmeGluTrrHleuGlu 352
4119913..... 4119913

352   nLeuLsYrrIlleGlyCySValIlleYsGlnHrLeuArGleuAsnPr 369
      :   :   :   :   :   :   :   :   :   :   :   :   :
4119912..GACGCTTCACGACGCGGTGGATCAAGACGATTCGCGGTGCTCGC 4119864

369   roValPrOglYglYPhearGValAlaLeuLysThrPheGluLeuAsn 385
      |||||:   :   :   :   :   :   :   :   :   :   :
4119863CGCGCGCTTACGCGCTGCTGTCGCGCGCACCACTAACATCGCGCGT 4119814

386   TyrGlnIllePrOlysglyTrrPaSnValIlleTrrSerTrrIecYs 402
      :   :   :   :   :   :   :   :   :   :   :   :   :
4119813TACCGCTGGAGGCTGGCACCAATCGGTGGTCCATCATCGCGATCAA 4119764

402   sAsPaValAlaAsPrllePheHrAsnLysGlnuPheAsnPrOaSpAr 419
      |||||:   :   :   :   :   :   :   :   :   :   :
4119763CGCGACGCCGAGGTATGAGACCCACACAGATTCGCGCCGAGCGGT 4119714

419   heIlleValPrOHsPrOgluAsPaLsArGPrHeserPheIlleProP 435
      :   :   :   :   :   :   :   :   :   :   :   :   :
4119713TCTCTCGTACCGGCGCGCAA.....ACCTATGCGTGGGTTCCGTC 4119673

436   GluYglYglYleuArGserCySValGlyLysGluPheAlaLysIlle 452
      |||||:   :   :   :   :   :   :   :   :   :   :
4119672GGCGCGCGCGGAGAACGCTGCTGGGTGGCACTTTCG....ATCGG 4119629

452   ulYsIllePheHrVal 457
      :   :   :   :   :   :   :   :   :   :   :   :   :
4119628AGAGTTGATCACCGTA 4119613

seq_name: /cgr2_6/prodata/2/lna/6B_COMB.seq:US-09-103-840A-1

seq_documentation_block:
  Sequence 1, Application US/09103840A
  Patent No. 6294328
  GENERAL INFORMATION:
  APPLICANT: FLEISCHMAN, Robert D.
  APPLICANT: WHITE, Owen R.
  APPLICANT: FRASER, Claire M.
  APPLICANT: VENTER, John C.
  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
  FILE REFERENCE: 24366-20007.00
  CURRENT APPLICATION NUMBER: US/09/103.840A
  CURRENT FILING DATE: 1998-06-24
  NUMBER OF SEQ ID NOS: 2
  SOFTWARE: Patentl Ver. 2.1
  SEQ ID NO 1
  LENGTH: 4411529
  TYPE: DNA

```



```

: TYPE: DNA
: ORGANISM: Mycobacterium tuberculosis
: FEATURE:
: OTHER INFORMATION: CDC 1551
: OTHER INFORMATION: "n" bases at various positions throughout the sequence
: OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Alignment_scores:
    Quality: 410.00      Length: 489
    Ratio: 1.507        Gaps: 23
    Percent Similarity: 55.624      Percent Identity: 29.243

Alignment block:
US-09-668-482-32 x US-09-103-840A-2 ..

Align seg 1/1 to: US-09-103-840A-2 from: 1 to: 4403765

37 AsparGSerCysAlaLeuProLeuProGlyThr..... 48
: : : : : : : : : : : : : : : : : : : : : : : : :
3415288GAGAGAAACATCGCGAACCG..CCGCCAGGAGCGGCGCTCAAGCCGT 3415334
49 .....MetGlyPheProPhePheGlyLutThrLeuMetVal 62
: : : : : : : : : : : : : : : : : : : : : : : : :
3415335CGTCGGTGAACGGCGGCGTCCGATCCTTGGCCACATGATGAGATTTGC 3415384
62 euGlnAArgArgPheLeuGlnMetLysArgArgLysTyrGlyPheIle 78
: : : : : : : : : : : : : : : : : : : : : : : : :
3415385GGGGGGGACCGGACATCTGATG.....TTCCGTG 3415413
79 TyrLysThrHis.....LeuPheGly.....ArgProThr 88
: : : : : : : : : : : : : : : : : : : : : : : : :
3415414TACCAAGACGAAGGGTCCGCTCGATTTCGGCGACTGCTGCTGCCGGG 3415463
88 rValArgValMetGlyAlaAspAsnValArgArgIleLeuGlyLysIle 105
: : : : : : : : : : : : : : : : : : : : : : : : :
3415464TGTGCGACAGCTGGGCGCCGACGCGCGGCGCATCTACTCCAAACCGCA 3415513
105 ISArgLeuValSerValHis...TyrProAlaSerValArgThrIleLeu 120
: : : : : : : : : : : : : : : : : : : : : : : : :
3415514ACCAAGACTACTCGACGAGGGGTGGTCCGATCGGCGCCCTTCTTC 3415563
121 GlyAlaGlyCysLeuSerAsnLeuHisAspSerSerHisLysGlnArgIly 137
: : : : : : : : : : : : : : : : : : : : : : : : :
3415564CAACGCCGCGC...CTCATGCTGCTCGACTTCGAAGACACATGTTCCACGG 3415610
137 sLysValIleMetGlnAlaPheSerArgIleAlaLeuGlnCysTyrValI 154
: : : : : : : : : : : : : : : : : : : : : : : : :
3415611ACGGATCTGACGAGGCGGTTCGTCGGGTCCAGCGCTCCGGCGCTACCTCG 3415660
154 euValIleAlaGlnGluIleValSerSerCysLeu...GluGlnTyrLeuSer 169
: : : : : : : : : : : : : : : : : : : : : : : : :
3415661AGCAGATGACAGGGGTCTCTCGCGGTGGTGGCCGACGACATGGGGTC... 3415707
170 CysGlyGluArgGlyLeuLeuValTyrProGluValLysArgLeuMetPhe 186
: : : : : : : : : : : : : : : : : : : : : : : : :
3415708GTCAACGACGACGCGCTTCTCTATCCGGCAGTAAGGCGCTACAGCT 3415757
186 eArgIleAlaMetArgIleLeuLeuGlyCysGluProGlyProAlaGlyG 203
: : : : : : : : : : : : : : : : : : : : : : : : :
3415758GACATCGCTCGATGCTTCATGCGGACGAAACCCGCGC..... 3415797
203 LysGlyLysAspGluGlnGlnLeuVal.....GluAlaPheGlu 215
: : : : : : : : : : : : : : : : : : : : : : : : :
3415798.....ACCGATTCACGAACTGGTACCAAGGTGACAAAGGCGTTACAG 3415839
216 GluMetThrArgAsn.....LeuPheSerLeuProIleAs 227
: : : : : : : : : : : : : : : : : : : : : : : : :
3415840ATTCACCAACCGCTCGCGCAACGGGTGATTCGGCAACCAACCGTCCCA... 3415884
227 pValProPheSerGlyLeuTyrArgGlyValLysAlaArgAsnLeuIleH 244
: : : : : : : : : : : : : : : : : : : : : : : : :

```



```

224 GAGAC,CGGCTGTGTGCTGCTCAGTGGCCAGCGTGGTGGCCAGCATTCCT 272
120 uG1yAlaG1yCysLeuSerAsnLeuHisAspSerHisLysGlnArgL 137
273 GGGATCTGGCTGCTCTTACCTGACAGACTCCCTGCAACAAGCAGACGA 322
137 yslYsValL1lMetGlnAlaPheSerArgGluAlaLeuGlnCysTyVal 153
323 AGAAGTGATTATGGGCGCTTCACCGCGAGGA,CTCGAATGCTACAGTG 371
154 LeuValL1leAlaGluLysValSerSerCysLeuGlnGlnTPrLeuSerC 170
372 CCGGATCGACCGAGAGAGTGGCGAGCTGGAGCGAGTGGCTGAGCTG 421
170 sG1yGluArgL1yLeuLeuValTyrProG1uValLysArgLeuMetPhe 187
422 CGGGAGAGGGCGCTCCCTGCTGTACCCGAGGTGAAGCGCTCATGTTCC 471
187 rG1leAlaMetArgL1leLeuLeuGlyCysGluProG1yProAlaG1yL 203
472 GAATCGCATCGCATCTACTGGGCTCGAAACCCCACTGGCGGCGAGC 521
204 G1yGluAspG1uGlnGlnValG1uAlaPheG1uGlnMetThrArgAs 220
522 GGGACATCCGACAGACACTGTGTGAGGCTTCGAGAAATGACCCGCA 571
220 nLeuPheSerLeuProIleAspValProPheSerC1yLeuTyrArgG1y 237
572 TCTCTCTCGCTGCCCATCGACGTGCCCTTCAGCGGCTGTACCGGGCA 621
237 alLysAlaArgAsnLeuL1eHisAlaArgL1leG1uGlnL1eArgAla 253
622 TGAAGCGCGGAGACCTATTTCACGCGCATCGACAGCAACAATTCGCC 671
254 LysL1leArgAlaG1uGlnAlaThrG1uProAspG1yG1yCysLysAsp 270
672 AAGATCTGGCGGCTGCGGGCATTCGAGCGGCGCAAGGCTCGAAAGAC 721
270 aLeuGlnLeuLeuL1leGlnHisSerTPrG1uArgG1yG1uArgLeu 287
722 GCTGACACTGTGTGATCGACACTGTGTGAGAGGAGGAGCGCTGAGCA 771
287 etGlnAlaLeuLysGlnSerSerThrG1uLeuLeuPheG1yG1yHisL 303
772 TGCAGGCACTAAAGCAATCTTCAACGAACTCCCTTTGGAGACACGAA 821
304 ThrThrAlaSerAlaAlaThrSerLeuL1eThrTyrLeuG1yLeuTyr 320
822 ACCAGCGCGCATGCGACCATCTCTGATCTACTGAGGCTCTACCC 871
320 oHisValLeuGlnLysValArgG1uGlnL1leYsSerLysG1yLeuLe 337
872 ACAATGTTCTCCAGAAAGTGGAGAAAGTGAAGAGTAAAGGTTACTTT 921
337 yslYsSerAsnGlnAspAsnLysLeuAspMetG1uTnTrLeuGlnLeu 353
922 GCAAGAGCAATCAAGACACAACTTGAACATGGAATTTTGGAACTTGA 971

seq_name: gb_est29:AL532445
seq_documentation_block:
LOCUS AL532445 893 bp mRNA EST 13-FEB-2001
DEFINITION AL532445 LTI_NFL001_NBC4 Homo sapiens cDNA clone CS0DM014YL13 5
prime, mRNA sequence.
ACCESSION AL532445
VERSION AL532445
KEYWORDS EST
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 893)
AUTHORS Li,W.B., Gruber,C., Jesse,J. and Polayes,D.

```

```

TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
source
1. 893
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DM014YL13"
/location="LTI_NFL001_NBC4"
/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="DH10B"
/note="Organ: Brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed
by Life Technologies. Contact : Feng Liang Life
Technologies, a division of Invitrogen 9800 Medical Center
Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
8371 Email : filiang@life.com url :
http://fulllength.invitrogen.com"
BASE COUNT 196 a 253 c 275 g 167 t 2 others
ORIGIN

```

```

alignment_scores:
Quality: 1255.00 Length: 307
Ratio: 4.498 Gaps: 4
Percent Similarity: 90.879 Percent Identity: 86.319

```

```

alignment_block:
US-09-668-482-32 x AL532445 ..

```

```

Align seg 1/1 to: AL532445 from: 1 to: 893

```

```

73 ArgLysTyrGlyPheL1eTyrLysThrHisLeuPheG1yArgProThrVa 89
2 AGGAATATCGGCTTCATCTACAGACGATCTGTGGCGGCCACCGCT 51
89 LArgValMetG1yAlaAspAsnValArgArgL1leLeuLeuGlnHisA 106
52 ACGGATGATGGCGCGGACATGTGGCGCATCTGTGCGAGA,SAAC 100
106 rGLeuValSerValHisTrpProAlaSerValArgThrL1leLeuG1yAla 122
101 GCGTGGTGTGCGTCCACTGGCGACGTCGCGGCACCATTCGTGGATCT 150
123 G1yCysLeuSerAsnLeuHisAspSerSerHisLysGlnArg, LysLysV 139
151 GCGTGCCTCTTACCTGACAGACTCCTCGCAAGACGCGCGAGGA... 196
139 alL1eMetGlnAlaPheSerArgGluAlaLeuGlnCysTyValLeuVal 155
197 .....AGCGCGAGGACATCGAATGCTACTAGTCCGGTG 229
156 L1leAlaGluLysValSerSerCysLeuGlnGlnTPrLeuSerCysG1yL 172
230 ATCACCGAGGAAGTGGCGAGCGAGCTGAGAGCATGCTGAGCTCGCGGA 279
172 uArgG1yLeuLeuValTyrProG1uValLysArgLeuMetPheArgL1le 189
280 GCGGCGCTCCTGCTTACCCCGAGGTGAAGCGCTCATGTTCCGAATCG 329
189 lAmetArgL1leLeuLeuG1yCysGluProG1yProAlaG1yG1yGlu 205
330 CCAATGGCAATCTACTGGGCTGCGAACCACATGCGGGCGAGCGGGAC 379
206 AspG1uGlnLeuValG1uAlaPheG1uGlnMetThrArgAsnLeuPhe 222

```

```

380 TCCGAGCAGCAGCTTGTGGAGGCGCTTCGAGGAATGACCCGCATCTCTT 429
222 eSerLeuProIleAspValProPheSerGlyLeuTyrArgGlyValLysA 239
|||||
430 CTCGCTGCCATTCAGACGTCCTTCAGCGGGCTGTACCGGGGCATGAAG 479
229 IaArgAsnLeuIleHisAlaArgIleGluGlnAsnIleArgAlaLysIle 255
480 CGCGGAACCTCATTACGCGGCATCGAGCAAGAACATTTCGGCGCAAGATC 529
256 ArgArgLeuGlnAlaThrGluProAspGlyGlyCysLysAspAlaLeuG1 272
|||||
530 TCGCGGCTGCCGGCATCCGAGCGGGCCAGGCTGCACAAAGCGCGCTGCA 579
272 nLeuLeuIleGluHisSerTrpGluArgGlyGluArgLeuAspMetGlnA 289
580 GCTGTTGATCGACGACTGCTGGAGAGGGAGACGCGCTGCATGCAAG 629
289 IaLeuLysGlnSerSerThrGluLeuLeu PheGlyGlyHisGluThrTh 305
|||||
630 CACTAAGCAATCTTCAACCGAAGCTCTTTTGGAGACACGAAACAC 679
305 rAlaSerAlaAlaThrSerLeuIleThrTyrLeuGlyLeuTyrProHisV 322
|||||
680 GGCGAGTCGACGACATCTGTGATCACTTACCTGGGCTTACCCACATG 729
322 aLeuGlnLysValArgGluGluIleLysSerLysGlyLeuLeuCysLys 338
|||||
730 TTCTCTCAAGAAAGTCGAGAGAGCTGAAGAGTTACTTTGCAAG 779
339 SerArgGlnAspAsnLysLeuAspMetGluThrLeuGlnGluIleLysTy 355
|||||
780 ACCAATTCAGACACACAGTGGACATGGAATTTTGGACA. CTTAAATA 828
355 rIleGlyCysValIleLysGluThrLeuArgLeuAsnProProValProg 372
|||||
829 CATCGGGTGTATTATTAAAGAGACTTCGACATGATCCCACT...TCAG 875
372 LyGlyPheArgValAla 377
|||||
876 GAGGCTTCGGGTGCT 892

seq_name: gb_est29:AL532444
seq_documentation_block:
LOCUS AL532444 1028 bp mRNA EST 13-FEB-2001
DEFINITION AL532444 LTI_NFL001_NBC4 Homo sapiens cDNA clone CSODM014YL13 3
ACCESSION AL532444
VERSION AL532444.1 GI:12795937
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1028)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
Location/Qualifiers
1. 1028
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CSODM014YL13"
/clone_lib="LTI_NFL001_NBC4"
/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="DH10B"
/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA

```

was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : <http://fulllength.invitrogen.com>

```

BASE COUNT      265 a      244 c      225 g      282 t      12 others
ORIGIN

alignment_scores:
    Quality: 1237.00      Length: 282
    Ratio: 4.739      Gaps: 3
Percent Similarity: 92.553      Percent Identity: 86.170

alignment_block:
US-09-668-482-32 x AL532444/rev ..

Align seg 1/1 to reverse of: AL532444 from: 1 to: 1028

217 MetThrArgAsnLeuPheSerLeuProIleAspValProPheSerGlyLe 233
|||||
1005 ATGACCCGA...ATCTCTCGCTCCCATCGACTGCCCT...TCAGGGGC 962

233 uTyrArgGlyValLysAlaArgAsnLeuIleHisAlaArgIleGluGlnA 250
|||||
961 TGTAACCGGATGTAGCGCGCAACCTTAATTAACGCGCGCATCGACGAGA 912

250 snIleArgAlaLysIleArgArgLeuGlnAlaThrGluProAspGlyGly 266
|||||
911 AMATAGCGGCCAAGATCTCGCGGCTCGCGGATCCGAGGGCGCAGAGC 862

267 Cys_LysAspAlaLeuGlnLeuLeuIleGluHisSerTrpGluArgGly 283
|||||
861 TGCAAAAGACCGCGCTGCAGCGTGTGATCGAGCACTCGTGAGAGGGAG 812

283 LuArgLeuAspMetGlnAlaLeuLysGlnSerThrGluLeuPhe 299
|||||
811 AGCGCTGGACATCGACGACATTAAGCAATCTTCACCGAATCTCTT 762

300 GlyGlyHisGluThrThrAlaSerAlaAlaThrSerLeuIleThrTyrLe 316
|||||
761 GGAGGACACGAAACACAGCGCCAGTGCCACATCTGTGATMACTTACT 712

316 uGlyLeuTyrProHisValLeuGlnLysValArgGluGluIleLysSerL 333
|||||
711 GGGGCTTACCCACATGTTTCCAGAAAGTCGAGAAAGAGCTGAAGAGTA 662

333 ysgLyLeuLeuCysLysSerAsnGlnAspAsnLysLeuAspMetGluThr 349
|||||
661 AGGGTTTACTTTGCAGAGCAATCAAGCAACAAGTTGGATGGAATGAAATT 612

350 LeuGluGlnLeuLysTyrIleGlyCysValIleLysGluThrLeuArgLe 366
|||||
611 TTGGACACACTTAATACATGGGTGTATTAAAGAGACCTTCGACT 562

366 uAsnProProValProGlyGlyPheArgValAlaLeuLysThrPheGlu 383
|||||
561 GAATCCCCCACTTCAGAGAGGTTTGGGTTGCTCGAAGACTTTGAAT 512

383 euAsnGlyTyrGlnIleProLysGlyTyrAsnValIleTyrSerIleCys 399
|||||
511 TAAATGGATACAGATTCCCAAGGCTGGAATGTATCTACAGATGTGT 462

400 AspThrHisAspValAlaAspIlePheThrAsnLysGluGluPheAsnPr 416
|||||
461 GATACTCATGATGTGGCGAGAGATCTTCACCAAGAGAGATTAATTC 412

416 oAspArgPheIleValProHisProGluAspAlaSerArgPheSerPheI 433
|||||

```

441 TGACGATTCATGCTGCTCACCACAGAGATGATCCAGTTCACGCTCA 362
 443 leProphgIyGlyLeuAArgSerCysValGlyLysGluPheAlaLys 449
 361 TTCATTGGAGAGGACCTTAGAGCTGTGTAGGCAAGAAATTTGGAAAA 312
 450 IleLeuLeuLysIlePhePheValGluLeuAlaArgHisCysAspTrpS1 466
 311 ATCTTCCAAATAATTACAGTGGAGCTGSCCAGCATGTSACTGCGM 262
 466 nLeuLeuAnGlyProProThrMetLysThSerProThrValTyrProV 483
 261 GCTTAAATGAGCYCTCCATGAAAWCYAGTCCACCGCTGTAATCC 212
 483 aLAspAnLeuProAlaArgPheThrTyrPheGlnGlyAspIle 497
 211 TGGACATCTCCCTCCAGATTCACCATTTCCATGGGGAATC 168
 seq_name: gb_est29:AL539667

seq_documentation_block:
 LOCUS AL539667 847 bp mRNA EST 16-FEB-2001
 DEFINITION AL539667 LTI_FL013_FBrn1 Homo sapiens cDNA clone CS0DF036Y04 3
 ACCESSION AL539667
 VERSION AL539667.1 GI:12869095
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 847)
 Li,W.B., Gruber,C., Jesse,J., and Polayes,D.
 Full-length cDNA libraries and normalization
 Unpublished (2001)
 COMMENT
 Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 Evry cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
 Location/Qualifiers
 1. 847

FEATURES
 source

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="CS0DF036Y04"
 /clone_lib="LTI_FL013_FBrn1"
 /dev_stage="pooled tissue from post conception fetuses (20
 week, 24 week and 26 week)."
 /lab_host="DH10B"
 /note="Organ: Petal brain; Vector: pCMVSPORT 6; 1st strand
 cDNA was primed with a NotI-oligo(dt) primer. Five prime
 end enriched, double-stranded cDNA was digested with Not I
 and enriched into the Not I and Eco RV sites of the
 pCMVSPORT 6 vector. Library was constructed by Life
 Technologies. Contact : Feng Liang Life Technologies, a
 division of Invitrogen 9800 Medical Center Drive Rockville
 , Maryland 20850, USA Fax : (1) 301 610 8371 Email :
 fliang@life-tech.com URL :
 http://fulllength.invitrogen.com"

BASE COUNT 197 a 190 c 204 g 229 t 27 others
 ORIGIN

alignment_scores:
 quality: 1001.00 Length: 231
 Ratio: 4.859 Gaps: 0
 Percent Similarity: 89.177 Percent Identity: 82.684

alignment_block:

US-09-668-482-32 x AL539667/rev ..

Align seq 1/1 to reverse of: AL539667 from: 1 to: 847

266 GlycylsAspAlaLeuGlnLeuLeuIleGlnHisSerTrpGluArg1 282

|||||
 846 GGTGCAAAAGACCGCTGCACTKTTGATGAGACACTCGTGAGAGGG 797
 282 ygluArgLeuAspMetGlnAlaLeuLysGlnSerSerThrGluLeuP 299
 796 AGAGCGGCTGAGCATGACAGCACTAAAGCAATCTTCAACGGAATCTCT 747
 299 heGlyGlyHisGluThrThrAlaSerAlaAlaThrSerLeuIleThrTyr 315
 746 TTGAGAGACAGAAACACGAGGCGAGTGACCAATCTGTGATCACTTAC 697
 316 LeuGlyLeuTyrProHisValLeuGlnLysValArgGlnGluIleLys 332
 696 CTGGGGCTCCACCCACACGCTCCCGAAGAGCCGAGAAAGACTGAGAG 647
 332 rLysGlyLeuLeuCysLysSerAsnGlnAspAsnLysLeuAspMetGlu 349
 646 TAAGGTTTACTTTGCAAGMCAATCAAGACAAGTTGGACATGAGAA 597
 349 hrLeuGlnGlnLeuLysTyrIleGlyCysValIleLysGluThrLeuArg 365
 596 TTTTGGAACTCMAATWCNCNGGGGTGTATTAAAGAGACCCCTCGA 547
 366 LeuAsnProProValProGlyLysPheArgValAlaLeuLysThrPheG1 382
 546 CTGAACCCCGCCGNCAGAGAGGTTNCGGNNGCTCCGAASMTCTTGA 497
 382 nLeuAnGlyTyrGlnIleProLysGlyTyrPAsnValIleTyrSerIleC 399
 496 ATCAATATGATCCCMANCCCMAGGCGTGAATGTATCTACACTACT 447
 399 yAspThrHisAspValAlaAspIlePheThrAsnLysGlnIlePheAsn 415
 446 GIGATACCATATGTGGCAGAGATCTTCCACCAAGAGAAATTTAAC 397
 416 ProAspArgPheIleValProHisProGluAspAlaSerArgPheSerP 432
 396 CCGACCATTCATCTGCTGCTCACCAGAGAGCAGCACCAGCCTCAGCC 347
 432 eLleProphgIyGlyLeuAArgSerCysValGlyLysGluPheAlaL 449
 346 CATCCATTTGGAGAGGCTTAGAGCTGTGTAGCAAAAGATTCGMA 297
 449 ySileLeuLeuLysIlePheThrValGluLeuAlaArgHisCysAspTrp 465
 296 AAATTCCTCCCAAAATATTACAGTGGAGCTGSCCAGCATTTGTACTSG 247
 466 GlnLeuLeuAnGlyProProThrMetLysThSerProThrValTyrPr 482
 246 CCGCBTCTAAATGAGACTCTACABGAAACCAAGCCACCGTGTATCC 197
 482 oValAspAnLeuProAlaArgPheThrTyrPheGlnLysAsp 496
 196 TGTGACATATCTCCCTGCAAGATTCACCATTCATGAGGAA 154
 seq_name: gb_est48:AW513600

seq_documentation_block:
 LOCUS AW513600 618 bp mRNA EST 03-MAR-2000
 DEFINITION x047D06.x1 NCI-CGAP-Ut1 Homo sapiens cDNA clone IMAGE:2707091 3'
 similar to SW:CP26_HUMAN 043174 CYTOCHROME P450 26 ;, mRNA
 sequence.
 ACCESSION AW513600
 VERSION AW513600.1 GI:7151678
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 618)
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index

JOURNAL
COMMENT

Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaabs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
image.llnl.gov/image/html/lresources.shtml

FEATURES

source
Possible reversed clone: similarity on wrong strand
Seq primer: -400p from Gibco
High quality sequence stop: 440.
Location/Qualifiers
1. 618
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2707091"
/clone_lib="NCI-CGAP_Ut1"
/tissue_type="well-differentiated endometrial
adenocarcinoma, 7 pooled tumors"
/lab_host="DH10B"
/note="Organ: uterus; Vector: PCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.75 Kb. Life Technologies catalog #: 11538-014"

BASE COUNT 111 a 191 c 199 g 115 t 2 others
ORIGIN

alignment_scores:
Quality: 982.00 Length: 206
Ratio: 4.935 Gaps: 0
Percent Similarity: 96.602 Percent Identity: 92.233

alignment_block:
US-09-668-482-32 x AW513600 ..

Align seg 1/1 to: AW513600 from: 1 to: 618

```

34 SerSerATGAspArgSerCysAlaLeuProleuProProGlyThrMetG1 50
1 AGCGGCCCGGAGCGAGTGTGCTCCATTCGCCCCGGGACTATGGG 50
50 yPheProPhePheGlyGluThrLeuGlnMetValLeuGlnArgArglySP 67
1 CTCTCCCTCTTTGGGGAACCTTCAGATGATGCTGACGCGAGGAAGT 100
67 heLeuGlnMetLeuArgArgGlyGlyThrLeuThrLeuThrHisLeu 83
101 TCCTTCAGATGAAGCGCAGAAATACGGCTTCATCTTAAGAAGCATCTG 150
84 PheGlyATGProthValArgValMetGlyAlaAspAsnValATGArg11 100
151 TTCGGGGCGGCCACCGTACGGGTGATGGCGCGGACATATGCGGGCAT 200
100 eleuLeuGlyGlnHisArgLeuValSerValHisTrpProAlaSerValA 117
201 CTTCGTCGAGAGCAGCGGCTGTGTCGTCACATGCGACGCTGCTGTC 250
117 rGThrTleLeuGlyAlaGlyCysLeuSerArgLeuHisAspSerSerHis 133
251 GCACCATTTCTGGATCTGCTCTCTTAACCTGCACGACCTCTCGAC 300
134 LysGlnArgLysLysValIleMetGlnAlaPheSerArgLysAlaLeuG1 150
301 AAGCAGCGGCAAGAAAGTATTTATGCGGGCTTCACCGCGAGGCACTGA 350
150 nGySTyValLeuValIleAlaGluGlnValSerSerCysLeuGluGlnT 167
351 ATGCTATGTCGTCGTCATCACCAGAGAGTGGGCGACGACCTGAGAGAGT 400

```

```

167 rpleuSerCysGlyGluArgGlyLeuLeuValTyrProGluValLysArg 183
401 GGCTGAGTCGCGCGCGAGCGGCTCTGCTGTACCCGAGGTGAAGCGC 450
184 LeuMetPheArgTleAlaMetArgIleLeuLeuGlyCysGlyProGlyPyr 200
451 CTCATTTTCGGAATCCGATCCGATCCGATCCGATCCGATCCGATCCG 500
200 oAlaGlyGlyGlyGluAspGluGlnGlnLeuValGluAlaPheGluGlu 217
501 GCGGCGCAGCGGAGACTCCGAGCAGCAGCTGTGAGGCGCTTCGAGGAA 550
217 ethrArgAsnLeuPheSerLeuProIleAspValProPheSerGlyLeu 233
551 TGACCGGCAATCTCTTCGCTGCTGCCATGACGATGCTTCAGCGGCTG 600
234 TyrArgGlyValLysAla 239
601 TACCGGGGATGAAGGCN 618
seq_name: gp_est72:BE236243

```

seq_documentation_block:
LOCUS BE236243 537 bp mRNA EST 10-JUL-2000
DEFINITION 143959 MARC 1Pig Sus scrofa cDNA 5', mRNA sequence.
ACCESSION BE236243
VERSION BE236243.1 GI:9020961
KEYWORDS EST.
SOURCE pig.
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
1 (bases 1 to 537)
Fahnenkrug,S.C., Frixling,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
Stone,R.T., Heaton,M.P., Grose,W.M., Bennett,G.A., Laegreid,W.W.
and Keele,J.W.
Design and use of two pooled tissue normalized cDNA libraries for
EST discovery in swine
Unpublished (2000)
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smithemall.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -mismatch 12 options.
PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCACAGTCAGCAGC
Plate: 85 row: B column: 21
Seq primer: ATTAGTGACACTATAG.
Location/Qualifiers
1. 537
/organism="Sus scrofa"
/db_xref="taxon:9823"
/clone_lib="MARC 1Pig"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: PCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from day 11, 13, 15, 20,
and 30 embryos."

JOURNAL
COMMENT
TITLE
EST discovery in swine
Unpublished (2000)
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smithemall.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -mismatch 12 options.
PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCACAGTCAGCAGC
Plate: 85 row: B column: 21
Seq primer: ATTAGTGACACTATAG.
Location/Qualifiers
1. 537
/organism="Sus scrofa"
/db_xref="taxon:9823"
/clone_lib="MARC 1Pig"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: PCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from day 11, 13, 15, 20,
and 30 embryos."

FEATURES
source
1. 537
/organism="Sus scrofa"
/db_xref="taxon:9823"
/clone_lib="MARC 1Pig"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: PCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from day 11, 13, 15, 20,
and 30 embryos."
BASE COUNT 96 a 163 c 174 g 104 t
ORIGIN
alignment_scores:
Quality: 900.00 Length: 179
Ratio: 5.114 Gaps: 0
Percent Similarity: 98.324 Percent Identity: 94.972


```

seq_name: gb_est89:BF533769
seq_documentation_block:
LOCUS BF533769 564 bp mRNA EST 11-DEC-2000
DEFINITION 602073982F1 NCL_CGAP_L19 Mus musculus cDNA clone IMAGE:4210893 5',
mRNA sequence.
ACCESSION BF533769
VERSION BF533769.1 GI:11621132
KEYWORDS house mouse,
EST.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1 (bases 1 to 564)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: c9apbs-r@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM9779 row: c column: 22
High quality sequence stop: 564.
FEATURES
Source Location/Qualifiers
1..364
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4210893"
/clone_lib="NCL_CGAP_L19"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: liver; Vector: pCMV-Sport6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.9 kb. Constructed by Life
Technologies. Note: this is a NCL_CGAP Library."
BASE COUNT 89 a 172 c 189 g 114 t
ORIGIN
alignment_scores:
Quality: 888.00 Length: 171
Ratio: 5.224 Gaps: 0
Percent Similarity: 99.415 Percent Identity: 99.415
alignment_block:
US-09-668-482-32 x BF533769 ..
Align seg 1/1 to: BF533769 from: 1 to: 564
1 MetGlyLeuProAlaLeuAlaSerAlaLeuGlyThrPheValLeuPr 17
|||||
50 ATGGGGCTCCGGCGCTGCTGCGCAGTGGCGCTGCTGCTGCTGCTGCC 99
|||||
17 GlnLeuLeuPheLeuAlaAlaLeuLysLeuTrpAspLeuTrpCysValS 34
|||||
100 GCTGCTGCTCTTCTGCGCGCTCAAGCTCTGGGACCTGTACTGTGTGA 149
|||||
34 GtSerArgAspArgSerCysAlaLeuProLeuProProGlyThrMetGly 50
|||||
150 GCGAGCGGCATCGACATCGCCCTCCCTTGCCTCCCGCGGTACATGGGC 199
|||||
51 PheProPheGlyGlyGluThrLeuGlnMetValLeuGlnArgArgLysPh 67
|||||
200 TTCACATCTTTTGGGAACATGACAGATGCTGCTCACGAGGAAGATT 249
|||||
67 eLeuGlnMetLysArgArgLysTrpGlyPheLeuTrpLysThrHisLeuP 84
|||||

```

```

250 TCTGCAGATGAAGCGCAGGAATAACGCTTCACTTCAACAGACCATCTGT 299
84 heGlyArgProThrValArgValMetGlyAlaAspAsnValArgArgGle 100
|||||
300 TTGGGGCGGCCACGATGCGGTGATGGCGCGGATATATGTGGCGCATC 349
|||||
101 LeuLeuGlyGluHisArgLeuValSerValHisTrpProAlaSerValAr 117
|||||
350 TTCTGTGGAGAGCACCGGTTGTGTGTGCTGCTGCTGCTGCTGCTGCG 399
|||||
117 gThrIleLeuGlyAlaGlyCysLeuSerAsnLeuHisAspSerSerHisL 134
|||||
400 CACCATCTCTGGCGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 449
|||||
134 ysgLArgLysLysValIleMetGlnAlaPheSerArgGluAlaLeuGln 150
|||||
450 AGCAGCGAAGAAAGAGTGATATGACAGCCCTTCACCGCGGAGGACCCG 499
|||||
151 CysTrpValLeuValIleAlaGluGluValSerSerCysLeuGlnTrp 167
|||||
500 TGCTACTGCGCCGATGCTGAGAGTGCAGATTTCTGAGACAGT 549
|||||
167 PleuSerCysGly 171
|||||
550 GCTTACGCTGCGGC 562
|||||
seq_name: gb_est71:BE189825
seq_documentation_block:
LOCUS BE189825 696 bp mRNA EST 07-AUG-2000
DEFINITION db61c05.x1 Wellcome CRC psk egg Xenopus laevis cDNA clone
IMAGE:3377480.5' similar to gb:gb1AF057566.11AF057566
laevis retinoid acid converting enzyme (XENOPUS);, mRNA sequence.
ACCESSION BE189825
VERSION BE189825.2 GI:9729548
KEYWORDS EST.
SOURCE African clawed frog.
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
Xenopodinae; Xenopus.
1 (bases 1 to 696)
Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L., Pape,D.,
Martin,J., Wylie,T., Underwood,K., Theising,B., Bowers,Y., Person
,B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R.,
Waterson,R. and Wilson,R.
Washu Xenopus EST project, 1999
Unpublished (1999)
On Jun 22, 2000 this sequence version replaced gi:8668718.
Other_ESTs: db61c05.x1
Contact: Sandy Clifton, Ph.D.
Washu Xenopus EST project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewatson.wustl.edu
Library constructed by N. Garrett, P. Lemaire, A.M. Zorn, and J.B.
Gurdon (Wellcome/CRC Institute). DNA Sequencing by: Washington
University Genome Sequencing Center
Clone distribution: Xenopus clones from this library are available
through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov
Seq primer: -40RP from Gibco
High quality sequence stop: 483.
FEATURES
Source Location/Qualifiers
1..696
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/clone="IMAGE:3377480"
/clone_lib="Wellcome CRC psk egg"
/tissue_type="egg"
/lab_host="DH10B (phage-resistant)"
/note="Vector: pBluescript SK-; Site_1: NotI; Site_2:

```

ECORI: cDNAs were oligo-dT primed and directionally cloned. Library was constructed by N. Garrett, P. Lemaire A.M. Zorn, and J.B. Gurdon (Wellcome/CRC Institute).
Note: This is a Xenopus Gene Collection (XGC) library."

Note: This is a Xenopus Gene Collection (XGC) library.

BASE COUNT	156 a	200 c	202 g	138 t
ORIGIN				

alignment_scores:	
Quality:	851.00
Ratio:	4.234
Percent Similarity:	85.532
	Gaps: 3
Percent Identity:	71.915

alignment_block:

Align seg 1/1 to: BE189825 from: 1 to: 696

43	ProteinProFrogLyThrMetGlyIleuProPhePhagIyLThrLeuG	59
44	
8	CCACTGGCCCCGGCACTAAGGGGCTGCCCTTCTGGGAGACTCTCA	57
59	mMetValLeuGlnArgArgLysPheLeuGlnMetLysArgArgLysTyrG	76
58	AATGGTCTCTCAAGAGCGCAAGTCTCTCCAAATGAAGCGTAGAAGTAG	107
76	lyPheIleTyrLysThrHisLeuPheGlyLysArgProThrValArgValMet	92
108	GTCCGCACTCTCAAAAGCAGCACTCTGTCTGGTAGCCCAAGCGTGGCGTCA	157
93	GlyAlaAspAsnValArgArgIleLeuLeuGlyGlnHisArgLeuValSe	109
158	GGCGCAGAGAAAGCTGGCAGAAATCCATATGGGGGAGCACAACGTGGTTC	207
109	rValHisTyrProAlaSerValArgThrIleLeuGlyAlaGlyCysLeus	120
208	GGTGCACATGGCCCGCTCGGTGGCGACGATCTCAAGGGCCGGCTGTCTGT	257
126	erAsnLeuHisAspSerSerHisLysGlnArgLysLysValIleMetGln	144
258	CCAACTGCACGACTGTGAGCACAAAGTACACCAAAAGAAATATTCACAA	307
143	AlaPheSerArgLysAlaLeuGlnCysTyrValIleValIleAlaGln	155
308	GCCTTCCTCCGAGATGCCCTCGCAATTAAGTGGCGGAGATGAAGAAAG	357
159	uValSerSerCysLeuGlnGlnTyrLeuSerCysGlyLysArgLysLeuL	176
358	GGTAGAAGCGTCTGTAAACCTGTGGCTGTGAGAGCGGCCCC...TGCGTGC	404
176	euAlaTyrProGluValAlaLysArgLeuMetPheArgIleAlaMetArgIle	192
405	TGATGTATCTCCGCACTCAAAAGCAGCTGATGTCTCCGATTCGCCATGAAGCTC	454
193	LeuLeuGlyCysGluProGlyProAlaGlyLysGlyLysAlaSpGlnGln	209
455	CTGCTGGCTGGCATGCC...CAGGCATGGACAGCCAACTAGAGGAGAC	501
209	nleuValGluAlaPheGlnGlnMetThrArgAsnLeuPheSerLeuProI	228
502	GGTGGCTGAAAGCTTCTGAGAGAAATACCCGAAATCTCTTCTCTGGCTTA	551
226	leAspValProPheSerArgLysLeuTyrArgGlyValLysAlaArgAsnLeu	242
552	TTCATGTGGCAGTATGGCGCCCTTACCGGGGCTTCGGGGCTAGAGACCTT	601
243	IleHisAlaArgIleGlnGlnAsnIleArgAlaLysAlaLysArgLeuG	259
602	ATTCATGCGCGGAATGATCAAAACATCGAAGAGAAGCTGTAAAG....	646
259	naIaThrGlnProAspGlyGlyCysLysAspAlaLeuGlnLeuLeuIleG	276
447GAACAGACGATAAATTCGGGGGAGGCCCTCAGCTGCTGATTG	689

```
276 1uhis 277
690 ACTAT 694
```

seq_name: gb_est4:AA239785

```
seq_documentation_block:
```

DEFINITION [mx80a03.r1](#) Soares mouse NML Mus musculus CDNA clone IMAGE:692620 5

mRNA sequence.

VERSION AA23978

SOURCE house mouse.

100

REFERENCE

AUTHORS

PTPIE

JOURNAL
COMMENT

Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMaGE Consortium (info@image.llnl.gov) for further information.
MGI:426180
Seq primer: -28m3 rev2 ET from Amersham
High quality sequence stop: 461.

Source

BASE CONT

ORIGIN

alignment_scores:	
quality:	845.00
ratio:	5.382
Percent Similarity:	99.367
alignment_block:	
length:	156
caps:	C
Percent Identity:	99.367

ORIGIN

[illegible]

```
alignment_block:
HE-00-6668-483-33 -- 11330706
```

Align seg 1/1 to: AA239785 from: 1 to: 474

324 GlnlySvalarggiugluilelyserlysglyleuleucyslyseras 340
 |||||
 1 CAGAAAGTCGAGAGAGATAAAGACCAAGGGCTTACTTTGCAGAGGCAA 50

```
340 nglaspasnllysleuaspmetglutThrLeuGlInleuLysTyrIleG 357
|||||
51 TCAGACACACAGATTAGACATGGAACCTTTGGACACGCTTAATACACTG 100
|||||
357 lYcysValIleLysGlutThrLeuArgLeuAsnProProValProGly 373
|||||
101 GGTGTGATTAAGAGACCTGCGATGGAATCCTCCGTTCCAGAGAGG 150
|||||
374 PheArgValAlaLeuLysThrPheGlLeuAsnGlyTyrGlnIleProLy 390
|||||
151 TTTCGGGTTGCTCTGAAGACTTTGAGCTGAATGATACAGATCCCA 200
|||||
390 sglTyrPasnValIleTyrSerIleCysAspThrHisAspValAlaAspI 407
|||||
201 GGGGTGATGTATTATTACAGTACTGTGACACCCACGATGGGACATA 250
|||||
407 lPheThrAsnLysGlutPheAsnProAspArgPheIleValProHis 423
|||||
251 TCTTCACTAACAGAGGAATTATCCCGACCGCTTATAGTGCCTCAT 300
|||||
424 ProGluAspAlaSerArgPheSerPheIleProPheGlyGlyLeuArg 440
|||||
301 CCAGAGGATGCTCCCGGTCCAGCTTCATTCATTGGAGAGGCGCTTCG 350
|||||
440 gSerCysValGlyLysGluPheAlaLysIleLeuLeuLysIlePheThrV 457
|||||
351 GAGCGTGTAGGCAAGAGATTGGCAAAATCTTCTTAAGATATTACAG 400
|||||
457 aGlutLeuAlaArgHisCysAspTrpGlnLeuLeuAsnGlyProProThr 473
|||||
401 TGGAGCTGGCTAAGGACCTGTATTGGCAGCTTCTAATAGGACCTCTACA 450
|||||
474 MetLysThrSerProThrValTyr 481
|||||
451 ATGAAGACACAGCCCACTGTGTAC 474

seq_name: gb_est85:BF236872

seq_documentation_block:
LOCUS BF236872 545 bp mRNA EST 14-NOV-2000
DEFINITION 602027980F1 NCI_CGAP_L19 Mus musculus cDNA clone IMAGE:4163202 5',
mRNA sequence.
ACCESSION BF236872
VERSION BF236872.1 GI:11150789
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 545)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/MLNL at:
http://image.llnl.gov
Plate: LHM9446 row: p column: 19
High quality sequence start: 2
High quality sequence stop: 545.
Location/Qualifiers
1. 545
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4163202"
/clone_lid="NCI_CGAP_L19"
```

```
/lab_host="DH10B (T1 phage-resistant)"
/notes="Organ: liver; Vector: pCMV-SpORF6; Site: 1; Note:
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.9 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."

BASE COUNT 87 a 165 c 184 g 109 t
ORIGIN

alignment_scores:
Quality: 805.00 Length: 173
Ratio: 4.820 Gaps: 4
Percent Similarity: 96.532 Percent Identity: 94.220

alignment_block:
US-09-668-482-32 x BF236872 ..

Align seg 1/1 to: BF236872 from: 1 to: 545

1 MetGlyLeuProAlaLeuLeuAlaSerAlaLeuCysThrPheValLeuPr 17
|||||
35 ATGGGGCTCCCGCGCTGCTGGCCAGTGCCTCTGCACCTTCGTGCTGCC 84
|||||
17 OleuLeuPheLeuAlaAlaLeuLysLeuTrpAspLeuTyrCysValS 34
|||||
85 GGTGCTGCTCTCTCGCGCGCGCTCAAGCTCGGACCTGTACTGTGTGA 134
|||||
34 eSerArgAspArgSerCysAlaLeuProLeuProGlyThrMetGly 50
|||||
135 GCAGCGCGCATGCGACCTCGCCCTCCCTTGGCCCCCGGTACCAAGGGC 184
|||||
51 PheProPhe.PheGlyGlutThrLeuGlnMetValLeuGlnArgArgLysP 67
|||||
185 TTCCCATTTCTTGAGGA...AACATGCAGATGGTGTTCACCGGAGGAAGT 231
|||||
67 heLeuGlnMetLysArgArgLysTyrGlyPheIleTyrLysThrHisLeu 83
|||||
232 TCTGCGAGATGAAGCGCAGGAATACGGCTTCATCTACAAAGACGATCG 281
|||||
84 PheGly.ArgProThrValArgValMetGlyAlaAspAsnValArgArgI 100
|||||
282 TTGGCGCGGCC...ACGGTCGGGTGATGGCGCGGATTAATGTGCGCGCA 328
|||||
100 lLeuLeuGlyGlnHisArgLeuValSerValHisTrpProAlaSerVal 116
|||||
329 TCTTGTGGGAGAGACCGGTTGGTGTGCGTGCATGCGCCGCGTGGTG 378
|||||
117 ArgThrIleLeuGlyAlaGlyCysLeuSerAsnLeuHisAspSerSerH 133
|||||
379 CCGACATGCTGGGCGCTGCTGCCCTCCCAACCTGCACGATTCCTCGCA 428
|||||
133 sLysGlnArgLysLysValIleMetGlnAlaPheSerArgGlyAlaLeuG 150
|||||
429 CAAGCAGCAAGAAAGAGTATTATGCAGGCTTCACCGCGAGGCACTCC 478
|||||
150 lncysTyrValLeuValIleAlaGluGluValSerCysLeuGluGln 166
|||||
479 AGTGCTACGTCGCCGTGATGCTGAGGAAGTCAAGCATTTGTGGAGCAG 528
|||||
167 TyrLeuSerCysGly 171
|||||
529 TGGCTAAGCTGCGCGC 543

seq_name: gb_est82:BF053367

seq_documentation_block:
LOCUS BF053367 646 bp mRNA EST 16-OCT-2000
DEFINITION 7178f12.x1 Soares NSF_P8_9M_OT_PA.P.S1 Homo sapiens cDNA clone
IMAGE:3392591 3' similar to SW:CR26_HUMAN O43174 CYTOCHROME P450 26
; mRNA sequence.
ACCESSION BF053367
VERSION BF053367.1 GI:10809263
KEYWORDS EST.
```

```

SOURCE          human.
ORGANISM        Homo sapiens
REFERENCE       Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS         1 (bases 1 to 646)
TITLE           NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
                National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
                Tumor Gene Index
JOURNAL         Unpublished (1997)
COMMENT         Contact: Robert Strausberg, Ph.D.
                Email: cgap@f-remail.nih.gov
                This clone is available royalty-free through LNL ; contact the
                IMAGE Consortium (info@image.llnl.gov) for further information.
                Seq primer: -40UP from GIBCO
                High quality sequence stop: 450.
                Location/Qualifiers
FEATURES
  source
    1..646
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone="IMAGE:3392591"
    /lab_host="DH10B"
    /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
    a modified polylinker; Site_1: Not I; Site_2: Eco RI;
    Equal amounts of plasmid DNA from five normalized
    libraries were mixed, and ss circles were made in vitro.
    Following HAP purification, this DNA was used as tracer in
    a subtractive hybridization reaction. The driver was from
    PCR-amplified cDNAs from pools of 5,000 clones made from
    the same 5 libraries. The pools consisted of the following
    libraries and cloneIDs: Soares NBHSF pool 1:
    309384-310919, 323208-325895 Soares NB2HP pool 1:
    145032-147335, 147720-148103, 148872-149255, 15002 -
    150407, 151176-152327 Soares NB2HP8-9W pool 1:
    758280-760583, 772104-774407 Soares NBHFA pool 1:
    304776-306311, 320136-322823, 326280-326663 Soares NBHOT
    pool 1: 723720-726407, 739080-740999 Subtraction by Bento
    Soares and M. Fatima Bonaldo."
BASE COUNT      202 a 129 c 120 g 194 t 1 others
ORIGIN
alignment_scores:
  Quality: 743.00      Length: 148
  Ratio: 5.160         Gaps: 0
  Percent Similarity: 97.297   Percent Identity: 91.216
alignment_block:
  US-09-668-482-32 x BF055367/rev ..
Align seg 1/1 to reverse of: BF055367 from: 1 to: 646
350 LeuGluGlnLeuLysTyrIleGlyGlyValIleLysGluThrLeuArgLe 366
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
643 TTGGCCCACTTAATTAATCTCGGCTGTATTAAGAGACCTTGACT 594
366 uAspProFovAlProGlyGlyPheArgValAlaLeuLysThrPheGluL 383
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
593 GAATCCCACTTCAGAGAGGTTTCGGGTCTTGAAGACTTTAAAT 544
383 euAsnGlyTyrGlnIleProLysGlyTyrPAsnValIleTyrSerIleCys 399
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
543 AAAATGATACCAAGATCCCAAGGCTGGAATGTATCTACAGTATCTGT 494
400 AspThrHisAspValAlaAspIlePheThrAsnLysGluGluPheAsnPr 416
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
493 GATACTCTATGATGGGAGAGATCTTCACCAACAAGAAATTTAAATCC 444
416 oAspArgPheIleValProHisProGluAspAlaSerArgPheSerPheI 433
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
443 TGACCGATTCATGCTGCTCACCCAGAGAGATGATCCAGGTTCACTTCA 394
433 leProPheGlyGlyLeuArgSerCysValGlyLysGluPheAlaLys 449

```

```

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
393 TTCATTTGGAGAGAGCCCTTAGAGCTGTGTAGGCAAAATTTGCAAAA 344
450 IleLeuLeuLysIlePhePheValGluLeuAlaArgHisCysAspTyrp1 466
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
343 ATTCTCTCAAAATATTATACGTAGTGAGCTGCAGGACATTTGAGTGGCA 294
466 lleuLeuAsnGlyProProThrMetLysThrSerProThrValTyrProV 483
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
293 GCTTCAATATGACCTCTACATGAAAAACAGTCCACCGGTATCTGT 244
483 aAspAsnLeuProAlaArgPheThrTyrPheGlnGlyAspIle 497
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
243 TGACACATCTCCCTGCAAGATTCACCCATTTCCATGGGGAATTC 200
seq_name: gb_est43:AM174347
seq_documentation_block:
  LOCUS      AM174347             720 bp      mRNA           16-NOV-1999
  DEFINITION f142a06.y1 Sugano Kawasaki zebrafish DRA Danio rerio cDNA clone
  2640274.5' similar to SW:CP26_BRARE P79739 CYTOCHROME P450 26 ;,
  mRNA sequence.
  ACCESSION  AM174347
  VERSION    AM174347.1 GI:6440295
  KEYWORDS   EST.
  SOURCE      zebrafish.
  ORGANISM   Danio rerio
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
              Cypriniformes; Cyprinidae; Rasbora; Danio.
REFERENCE    1 (bases 1 to 720)
AUTHORS      Sugano,S., Kawakami,K., Johnson,S., Li,F., Marra,M., Eddy,S.,
              Hillier,L., Clifton,S., Allen,M., Gibbons,M., Jost,S., Kueba,T.,
              Martin,J., Pape,D., Steptoe,M., Underwood,K., Thelsting,B., Rilter
              E., Bowers,Y., Wylie,T., Waterson,R. and Wilson,R.
              Washu zebrafish EST Project 1999
              Unpublished (1999)
JOURNAL      Other ESTs: f142a06.x1
COMMENT      Contact: S.L. Johnson
              Washington University School of Medicine
              4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
              Tel: 314 286 1800
              Fax: 314 286 1810
              Email: est@watson.wustl.edu
              Library constructed by Dr. Sumio Sugano and Dr. Koichi Kawakami DNA
              Sequencing by: Washington University Genome Sequencing Center
              Seq primer: T3 ET from Amersham
              High quality sequence stop: 478.
              Location/Qualifiers
FEATURES
  source
    1..720
    /organism="Danio rerio"
    /strain="AB"
    /db_xref="taxon:7955"
    /clone="2640274"
    /clone_lib="Sugano Kawasaki zebrafish DRA"
    /sex="mixed (one male and one female, including
    unfertilized eggs)"
    /dev_stage="adult"
    /lab_host="DH10B (phage resistant)"
    /note="Vector: pME18S-FL3; Site_1: DraIII (CACTGTGTG);
    Site_2: DraIII (CACCATGTG); 1st strand cDNA was primed
    with an oligo(dT) primer [ATGTGGCTTTTCTTTTCTTTT];
    double-stranded cDNA was ligated to a DraIII adaptor
    [TGTGGCTTACTGTG], digested and cloned into distinct DraIII
    sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site
    CACCATGTG). XhoI should be used to isolate the cDNA
    insert. Size selection was performed to exclude fragments
    <1.5kb. Library constructed by Dr. Sumio Sugano
    (University of Tokyo Institute of Medical Science) and
    kindly donated by Dr. Koichi Kawakami. Custom primers for
    sequencing: 5' end primer CTCTGTCTTAAAGCTGGC and 3' end
    primer CGACCTGACGCTCGACACA."
BASE COUNT    172 a 176 c 196 g 176 t

```



```

:::|||||:::|||||:::|||||:::|||||:::|||||
212 TTGCGGTCATTTGGAGAAACCTCCAGCTGATCTCCAGAGAGAGAGTT 261
67 eLeuGlmEtlYsArGArGlyTYrGlyPheIleTYrLYsThrHisLeuP 84
|||||:::|||||:::|||||:::|||||:::|||||
262 TCTGGGCAATGAACGGCGAATAACGGGTGATCTACAAGACGCACTCT 311
84 heGlyArGProThrValArGValImetGlyAlaAspAsnValArGArgIle 100
|||||:::|||||:::|||||:::|||||:::|||||
312 TCGGGAACCCGACTGTCAGGGTGATGGAGCTGATTAATGTGAGCGAGATT 361
101 LeuLeuGlyGIuHisArGLeuValSerValHisTrpProAlaSerValAr 117
|||||:::|||||:::|||||:::|||||:::|||||
362 CTGCTGGGCGAACAACAGCTGTGTCTCTTCAGTGGCCAGCATCAGTGAG 411
117 gThrIleLeuGlyAlaGlyCYsLeuSerAsnLeuHisAspSerHisL 134
|||||:::|||||:::|||||:::|||||:::|||||
412 AACCATCTCTGGGCTCTGACACCTCTCCAAATGTCATGAGATTCAACACA 461
134 ySGInArGLySLyValIleMetGlnAlaPheSerArGLuAlaLeuGln 150
|||||:::|||||:::|||||:::|||||:::|||||
462 AAAACAAAGAAAAAGGCCATTATGAGGGGCTCTCTCGAGATGCTCTGGAG 511
151 CysTYrValleuValIleAlaGluGluValSerSerCYsLeuGluGlnTr 167
|||||:::|||||:::|||||:::|||||:::|||||
512 CACTACATTTCCCGTGTATCCAGCAGAGAGGTGAAGAGCGCCATACAGGAATG 561
167 pLeu.....SerCYsGlyGluArGLyLeuLeuValTYrProGluV 181
|||||:::|||||:::|||||:::|||||:::|||||
562 GCTGCATAAAGACTCTCTGC.....GTGCTGGTTTATCCAGAAA 599
181 alLyArGLeuMetPheArgIleAlaMetArgIleLeuLeu 194
:::|||||:::|||||:::|||||:::|||||:::|||||
600 TGAAGAAACTCATGTTTGGATAGCTATGAGAAATCTGCTT 640
```

